Establishment of Heterochromatin at rRNA Genes Is Required for Embryonic Stem Cell Differentiation

Dissertation

zur

Erlangung der naturwissenschaftlichen Doktorwürde
(Dr. sc. nat.)

vorgelegt der
Mathematisch-naturwissenschaftlichen Fakultät

der
Universität Zürich

von

Sergio Leone

aus Italien

Promotionskommission

PD Dr. Raffaella Santoro (Leitung der Dissertation)

Prof. Dr. Dr. Michael O. Hottiger (Vorsitz der Dissertation)

Prof. Dr. Ulrike Kutay

Prof. Dr. Christian Lehner

Dr. Davide Gabellini

Zürich, 2017

"Faber est suae quisque fortunae"

Appius Claudius Caecus,
Roman consul

"There will always be rocks in the road ahead of us. They will be stumbling blocks or stepping-stones; it all depends on how you use them."

Friedrich Nietzsche,
German philosopher

Summary

The nucleolus, the most prominent substructure within the nucleus, is the compartment where transcription of hundreds of ribosomal RNA (rRNA) genes, rRNA processing, and ribosome subunit assembly takes place. Even though we have been aware of the nucleolus for 180 years, our knowledge of nucleolar functions is limited. Indeed, in the past 40-50 years, the major aim of research on nucleolus was to define the components and key steps of ribosome synthesis and it was not suspected that it could be involved in many other processes.

In differentiated cells a fraction of the ca. 400 copies of rRNA genes is transcriptionally silent and organized in compact heterochromatic structures, characterized by DNA methylation and repressive histone marks. The silent heterochromatic state of rRNA genes is stably inherited across cell divisions and is not affected even in case of high metabolic activities of the cell. Until now, the role of silent heterochromatic rRNA genes is not fully understood.

TIP5, a component of NoRC complex, together with the long non-coding (lnc)RNA pRNA establishes heterochromatin at rRNA genes. TIP5 binds to rRNA genes via pRNA and recruits DNA methyltransferases and histone modifier enzymes to establish silencing. pRNA, which originates from the processing of the intergenic spacer (IGS)-rRNA, is necessary for TIP5 recruitment to rRNA genes.

In embryonic stem cells (ESC) a large portion of the genome is euchromatic, a structure that well reflects the plasticity and transcriptional permissiveness of ESC genome that has to have the ability to enter any distinct transcriptional programs for lineage specification. Upon differentiation, large-scale genome silencing takes place and a broad part of ESC chromatin undergoes structural remodeling toward a highly condensed heterochromatic and transcriptionally repressed form. Recent results have shown that in ESCs all rRNA genes are euchromatic and acquire heterochromatic marks only upon differentiation. In ESCs, TIP5 does not repress rRNA genes due to impairment of the precursor lncRNA IGS-rRNA processing into the mature pRNA, which is the functional lncRNA required for recruitment of TIP5 to rRNA genes and establishment of rRNA gene silencing. Importantly, addition of mature pRNA in ESCs was not only sufficient to recruit TIP5 to nucleoli and silence rRNA genes but also induced the establishment of highly condensed chromatin structures outside of

the nucleolus, resembling the genome organization that characterizes differentiated cells. These results indicated that the nucleolus is not only the cellular compartment where ribosomes are produced but it is also able to produce heterochromatin, affecting the genome architecture of the rest of the nucleus.

In this work, we aimed to (1) identify which factors are implicated in IGS-rRNA processing and (2) determine the functional significance for the formation of heterochromatin at rRNA genes during ESC differentiation. Using a screening for IGS-rRNA-binding proteins we identified the RNA helicase DHX9 as key regulator of IGS-rRNA processing. DHX9-mediated pRNA production is required to guide TIP5 to rRNA genes and to establish heterochromatin. In ESCs DHX9 is not associated with rRNA genes and only upon ESC differentiation it localizes within nucleoli and binds to rRNA genes. Depletion of DHX9 in ESCs does not affect self-renewal and important traits of pluripotency such as expression of pluripotency factors. However, ESCs depleted of DHX9 are unable to differentiate and undergo cell death. Strikingly, this phenotype can be reverted by adding mature pRNA before induction of differentiation of ESCs depleted of DHX9.

Taken together these results highlight the role of lncRNA in the regulation of chromatin and epigenetic states and suggest that lncRNA processing represents an additional level of lncRNA regulation, which modulates distinct features of the same lncRNA. Moreover, they demonstrate that the state of nucleolar chromatin at rRNA genes is part of the regulatory network that controls exit from pluripotency and initiation of differentiation pathways.

Zusammenfassung

Der Nukleolus ist die prominenteste Struktur innerhalb des Zellkerns und ist der Ort an dem die Transkription hunderter ribosomaler RNA (rRNA) Gene, rRNA Verarbeitung und Ribosom-Zusammensetzung stattfindet. Obwohl Nukleoli bereits vor 180 Jahren beschrieben worden sind, ist unser Wissen bezüglich ihrer Funktion limitiert. Tatsächlich war das Hauptziel der Forschung der vergangenen 40-50 Jahren lediglich auf die Charakterisierung der Komponenten und zentralen Schritte der Ribosomsynthese begrenzt, ohne dabei zu ahnen, in welchen Prozessen Nukleoli ebenfalls entscheidende Rollen spielen.

In differenzierten Zellen wird ein wesentlicher Anteil der ca. 400 rRNA Genkopien ausgeschaltet und in kompaktem Heterochromatin organisiert, welches durch DNA CpG Methylierung und repressive Histonmodifikationen charakterisiert ist. Der heterochromatische Genzustand wird dabei auch über die Zellteilung hinaus an die nächste Zellgeneration vererbt und bleibt auch von der metabolischen Zellaktivität unberührt. Die Rolle der inaktiven heterochromatischen rRNA Gene ist jedoch noch nicht vollständig verstanden.

TIP5, eine Untereinheit des NoRC Komplexes, etabliert in Zusammenarbeit mit einer langen nicht-kodierenden RNA (lncRNA), pRNA, Heterochromatin in rRNA Genen. TIP5 bindet die rRNA Gene mit Hilfe der pRNA und rekrutiert DNA Methyltransferasen und histonmodifizierende Enzyme, die zur Abschaltung der Gene führen. Die pRNA, die aus der Verarbeitung der sogenannten intergenic spacer (IGS)-rRNA stammt, ist dabei für die TIP5-Rekrutierung zu den rRNA Genen unentbehrlich.

In embryonalen Stammzellen (ESC) befindet sich die Mehrheit des Genoms in einem euchromatischen Zustand, welches den Zellen erlaubt, einen hohen genomischen Grad an Plastizität und Transkription zu gewährleisten, welches ihnen ermöglicht, jedes beliebige zellinienspezifische Transkriptionsprogramm anzunehmen. Auf Initiation des Differenzierungsvorgangs finden weiträumige Umstrukturierungen statt, bei denen weite Teile des Genoms ausgeschaltet werden, indem die Transkription verhindert und Heterochromatin etabliert wird. Jüngste Ergebnisse haben gezeigt, dass sämtliche rRNA Gene in ESCs in euchromatischen Zustand vorliegen und erst

der Differenzierungsvorgang führt zu einem Übergang in einen heterochromatischen Zustand. In ESCs unterdrückt TIP5 die RNA Gene nicht aufgrund der mangelnden Verarbeitung der IGS-rRNA in pRNA, die als funktionsfähige lncRNA für die TIP5-Rekrutierng zu den rRNA Genen nötig ist und es TIP5 erlaubt, dort Heterochromatin zu etablieren. Es ist jedoch hervorzuheben, dass die Hinzufügung von pRNA zu den ESCs nicht nur ausreichte TIP5 zu den Nukleoli zu rekrutieren und ribosomale Gene auszuschalten, sondern auch heterochromatische Strukturen außerhalb der Nukleoli zu initiieren, ein Vorgang, der differenzierte Zellen charakterisiert. Diese Ergebnisse deuten darauf hin, dass der Nukleolus nicht nur für die Ribosomsynthese entscheidend ist, sondern auch Heterochromatin erzeugt, welches die genomische Architektur des gesamten Nukleus beeinflusst.

In der vorliegenden Arbeit haben wir uns zum Ziel gesetzt (1) die Faktoren zu identifizieren, die für das Prozessieren der IGS-rRNA verantwortlich sind und (2) die funktionelle Signifikanz der Heterochromatin-Formation in rRNA Genen während der Differenzierung zu ermitteln. In einem Screening zur Identifizierung von IGS-rRNA-bindenden Faktoren gelang uns die Bestimmung RNA Helikase DHX9 als einen zentralen Regulator der IGS-rRNA Verarbeitung. Die DHX9-vermittelte pRNA Produktion ist nötig für die gezielte Führung von TIP5 zu den ribosomalen Genen und der Heterochromatin Erzeugung. In ESCs ist DHX9 nicht mit rRNA Genen assoziiert und nur durch den Differenzierungsvorgang wird DHX9 in den Nukleoli lokalisiert, in denen es dann an die rRNA Gene bindet. Die Erschöpfung von DHX9 in Stammzellen führt zu keinem Verlust ihrer Selbsterneuerungsfähigkeit und ändert ebenfalls nichts an Pluripotenzmerkmalen. ESCs in denen DHX9 erschöpft wurde, sind jedoch nicht mehr im Stande zu differenzieren und weisen hohen Zelltod vor. Auffällig ist, dass dieser Phänotyp durch die Zugabe von pRNA vor der Zelldifferenzierung rückgängig gemacht werden kann, obwohl DHX9 in den Zellen fehlt.

Zusammengefasst heben diese Ergebnisse die Rolle von lncRNA in der Regulierung von Chromatin und epigenetischen Zuständen hervor. Außerdem deuten diese Ergebnisse darauf hin, dass das Prozessieren von lncRNA als weiteres Regulierungslevel der Eigenschaften der selben lncRNA dienen könnte. Des weiteren demonstrieren sie, das der Zustand des nukleolären Chromatins Teil des regulatorischen Netzwerks ist, welches den Austritt aus der Pluripotenz kontrolliert und den Differenzierungspfad initiiert.

Abbreviations

5hmC	5-hydroxymethylcytosine
5mC	5-methylcytosine
bp	base pair
CGI	CpG island
DC	dosage compensation
DNMT	DNA methyltransferase
dsRNA	double stranded RNA
ESC	embryonic stem cell
HAT	histone acetyl transferase
HAT	high affinity site
HDAC	histone de-acetylase
HDM	histone de-methylase
HMT	histone methyl transferase
ICM	inner cell mass
IGS-rRNA	intergenic spacer ribosomal RNA
KTM	lysine methyl transferase
lincRNA	long intergenic non coding RNA
lncRNA	long non coding RNA
miRNA	micro RNA
MRE	MSL recognition element
MSL	male-specific lethal
ncRNA	non coding RNA
NSL	non-specific lethal
NOR	nucleolus organizing region
NoRC	nuclelar remodeling complex
nt	nucleotides
pRNA	promoter RNA
piRNA	PIWI interacting RNA
PTM	post-translational modification
RISC	RNA induced silencing complex

RNP	ribonucleoprotein
RTM	arginine methyl transferase
rDNA	ribosomal DNA
roX	RNA on X (chromosome)
rRNA	ribosomal RNA
SAM	S-adenosyl-L-methionine
siRNA	small interfering RNA
snRNA	small nuclear RNA
snoRNA	small nucleolar RNA
tiRNA	transcription initiation RNA
TIP5	TTF1 interacting protein 5
TSS	transcription start site
TTF1	transcription termination factor 1
UBF	upstream binding factor
UCE	upstream control element
XCI	X chromosome inactivation
Xi	inactive X chromosome
Xic	X inactivation center
Xist	X inactive specific transcript

Table of Contents

1 Introduction

1.1 Chromatin structure

Eukaryotic genomes are organized in linear chromosomes that are enclosed in the nucleus of the cells. Each chromosome is composed of genomic DNA and structural proteins called histones. The genomic DNA is wrapped around histone proteins forming a fundamental structure called nucleosome, which is assembled into a higher-order structure known as chromatin (**Figure 1**). Nucleosomes are composed of two copies of each core histone, namely H2A, H2B, H3 and H4, forming an octameric structure that is able to bind 147 bp of DNA. The DNA is coiled in a left-handed super-helical conformation that turns 1.67 times around the histone octamer (Ramakrishnan, 1997; Richmond and Davey, 2003). Linker DNA of variable lengths connects adjacent nucleosomes resulting in the so-called 10 nm or "beads-on-a-string" fiber (Olins and Olins, 1974). A fifth histone protein, the histone H1, is not included in the core nucleosome structure but can bind 20 bp of the linker DNA and increase the DNA associated to the histone octamer to 166 bp, corresponding to two super helical turns (Ramakrishnan, 1997; Thoma et al., 1979; Zhou et al., 2015).

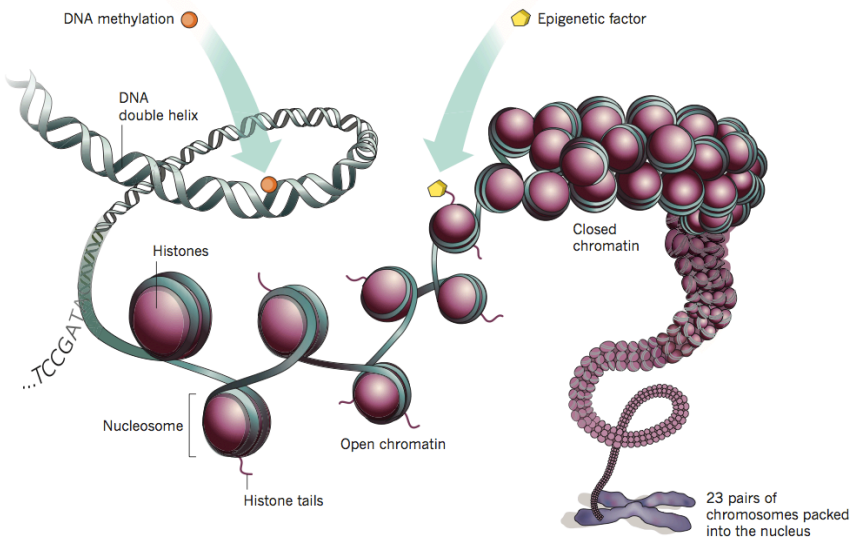


Figure 1. Structure of chromatin. DNA molecule wrapped around nucleosomes gives rise to a higher order structure allowing for compaction into the characteristic chromatin fiber and chromosomes. DNA methylation influences which genes are expressed, and other epigenetic factors (e.g. histone modifications) determine the compaction status of chromatin. Hence, both epigenetic marks control the transcriptional and compaction state of DNA without changing the underlying DNA sequence. From (Marx, 2012)

The 10 nm fiber represents the first layer of chromatin organization. A further step of compaction is represented by the 30 nm fiber that has been described under physiological salt concentration *in vitro*. The exact structure of the 30 nm fiber is still under debate but two models arose in the past describing a one-start solenoidal helix structure (Finch and Klug, 1976; (Robinson et al., 2006) or a two-start helix, in which nucleosomes are assembled in a zigzag ribbon that twists or supercoils (Dorigo et al., 2004; Finch and Klug, 1976; Schalch et al., 2005; Woodcock et al., 1984). It cannot be anyway excluded that the two models actually co-exist *in vivo*, or even that interphase chromosomes are assembled through long-range interactions of extended 10 nm fibers that form irregular interdigitated polymers that do not require the assembly of a regular 30 nm fiber (Maeshima et al., 2014; Maeshima et al., 2016). This last model is also strongly supported by recent experiments using electron spectroscopic imaging combined with electron tomography, which describe how the entire genome is organized in 10 nm fibers that are more or less compacted in irregular structures (Fussner et al., 2012). Also biochemical approaches based on proximity ligation procedure like Hi-C analysis, which explore the three-dimensional architecture of the whole genome, excluded further hierarchical order of chromatin organization than the 10 nm fiber (Lieberman-Aiden et al., 2009).

1.1.1 Euchromatin and heterochromatin

Chromatin has been structurally and functionally classified as euchromatin, which is transcriptionally active, structurally loose and characterized by the presence of active histone marks (see 1.1.2), and as heterochromatin that by opposite is poorly transcribed, is more compact and labeled by repressive histone marks (Felsenfeld and Groudine, 2003). However the distinction between the transcriptionally active and structurally loose euchromatin and the inactive and more compact heterochromatin most likely relies on the polymer melt model (Razin and Gavrilov, 2014) describing a disordered compaction of the 10 nm fiber rather than a defined distinction between respectively a 10 nm and a 30 nm fiber (Felsenfeld and Groudine, 2003).

Chromatin is not only composed of DNA and histone proteins, but it is actually a melting pot of regulatory proteins, which are more or less tightly bound to histones or DNA, and RNA, which exert a regulatory and also structural function in the regulation of chromatin architecture. Despite this complexity, chromatin is indeed a

very dynamic structure that contributes to the regulation of gene expression according to its degree of compaction. The heterochromatic state is more compact and for this reason less accessible to the transcriptional machinery compared to the euchromatic state. The degree of compaction is regulated by a cooperation of histone post-translational modifications (PTMs), DNA methylation and chromatin remodeling factors. Together with DNA methylation, histone PTMs are at the basis of the epigenetic inheritance of gene function. Epigenetics is by definition "the study of mitotically and/or meiotically heritable changes in gene function that cannot be explained by changes in DNA sequence" (Russo et al., 1996). This means that gene function does not depend exclusively on the DNA sequence that is inherited from a mother cell to a daughter cell, but it depends also on the chromatin context in which the gene is included and that can be as well inherited during cell division.

1.1.2 Histone post-translational modifications

Histones are alkaline proteins containing three alpha helices connected by two unstructured loops in a globular motif that allows protein dimerization, assembly of the octamer and interaction with DNA (Alva et al., 2007; Luger et al., 1997). Besides the globular histone fold domain, histones possess N-terminal unstructured "tails" that protrude from the histone octamer and can form additional contacts with DNA and chromatin associated proteins. Histones can be post-translational modified both at the globular domain (Tessarz and Kouzarides, 2014) and at the N-terminal tail (Kouzarides, 2007). PTMs can occur at different residues such as lysine (K), arginine (R), threonine (T), tyrosine (Y), serine (S), methionine (M), proline (P) and glutamate (E). Most common PTMs are characterized by methylation (me) of lysine and arginine; acetylation (ac), ubiquitylation (ub) and sumoylation (su) of lysine residues; phosphorylation (ph) of serine, threonine and tyrosine (Basnet et al., 2014); and mono- or poly-ADP-ribosylation of glutamate (E) and arginine (R) residues (Hassa et al., 2006). Less abundant and less characterized PTMs include citrullination of arginine (Cuthbert et al., 2004; Hagiwara et al., 2002), acylation of lysine besides acetylation (Rousseaux and Khochbin, 2015), oxidation of methionine (Luense et al., 2016), glycosylation of serine and threonine (Sakabe et al., 2010) and proline isomerization (Nelson et al., 2006). All these PTMs cause changes in the biophysical and biochemical properties of histones that ultimately result in the local alteration of

chromatin structure and protein association. In the past years the complex array of histone modifications and their functional meaning have been simplified in relation to the transcriptional status of the associated genes, discriminating between active and repressive histone marks (**Figure 2**). This lead to the concept of "histone code" described by Strahl and Allis (Strahl and Allis, 2000). Nowadays this idea has evolved due to the fact that several histone modifications revealed to have more complex effects that not only correlates with the transcriptional status of the associated genes but are also involved in the regulation of other nuclear processes like DNA replication, DNA damage response and nuclear architecture.

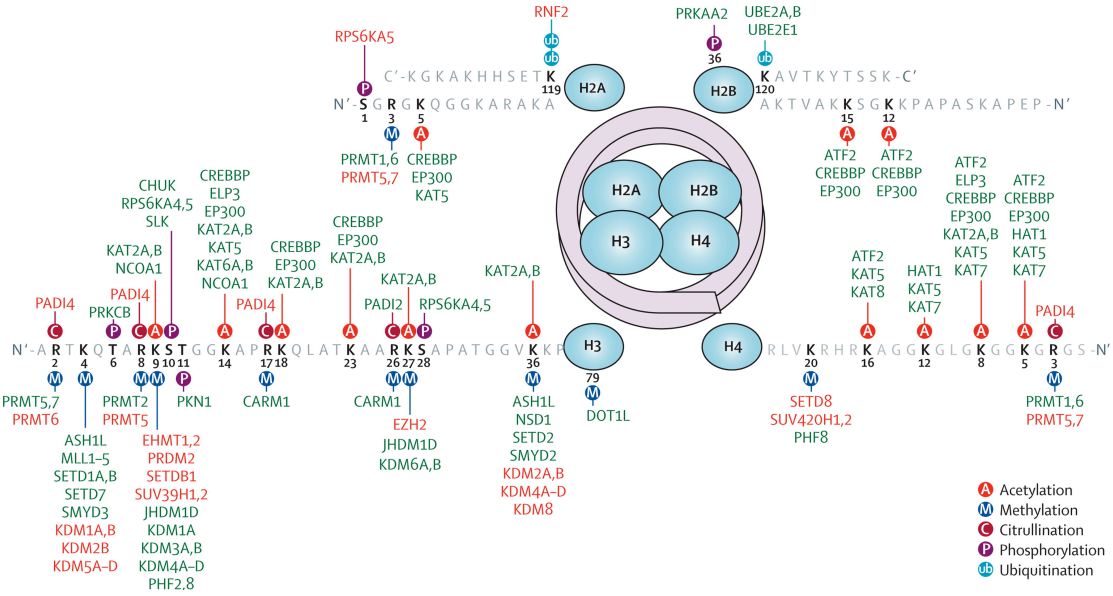


Figure 2. Major histone PTMs. Most common PTMs of lysine (K), arginine (R), serine (S), and threonine (T) residues are shown with their respective modifying enzymes (i.e. "writers" or "erasers"). Lysine residues can be monomethylated, dimethylated, or trimethylated. Enzymes shown in green are associated with transcriptional activation, and enzymes shown in red are associated with transcriptional repression. From (Huynh and Casaccia, 2013)

For this reason the concept of "histone code" has been gradually replaced by the more suitable concept of "histone language" (Rothbart and Strahl, 2014). Nevertheless several histone PTMs have been well characterized in relation to their impact on the regulation of gene expression. Therefore for instance H3K4me1/3, H3K9ac, H3K14ac, H3K27ac H3K36me3, H4K20me are considered "active" histone marks because mainly associated with open euchromatic regions of the genome containing actively transcribed genes whereas H3K9me2/3, H3K27me2/3, H3K79me3, H2BK5me3 are "repressive" histone marks associated with transcriptionally inactive regions of the genome. All these PTMs and many others are added by "writer" and

deleted by “eraser” enzymes, and their biological meaning is due to “reader” proteins that recognize the modification and trigger a biological output. Histone methylation is catalyzed by histone methyltransferases (HMTs). Both lysine methyltransferases (KMTs) and arginine methyltransferases (RMTs) use S-adenosyl-L-methionine (SAM) as methyl group donor (Black et al., 2012; Morales et al., 2016). Most of the KMTs (with the only exception of KTM4/Dot1L, (Okada et al., 2005)) share the same catalytic SET domain first described in Su(var)3-9, Enhancer of Zeste and Trithorax proteins (Jenuwein, 2006). Lysine methylation can occur as mono-, di- or tri-methylation and a specific enzyme commonly mediates each modification. For instance KMT1C (G9a) and KMT1D (GLP) are responsible for mono- and di-methylation of H3K9 while tri-methylation is catalyzed by KMT1A/B (SUV39H1/2) and KMT1C. H3K27me3 is catalyzed by E(z) in *Drosophila* and by the homologous proteins EZH1/2 in mammals (Grossniklaus and Paro, 2014). Histone demethylases (HDM) represent the erasers of histone methylation and belong to two families, LSD that use flavin-adenine-dinucleotide (FAD) as cofactor and JmjC that instead use α-chetoglutarate, molecular oxygen and Fe(II) to oxidize and remove the methyl group (Black et al., 2012). Methylated lysine is “read” by proteins containing chromodomains, PHD fingers, WD40 domains or ankyrin repeats (Glatt et al., 2011), while methylated arginine is recognized by proteins containing Tudor domains (Chen et al., 2011).

Histone acetylation is mediated by histone acetyltransferases (HATs) belonging to five families: HAT1, Gcn5/PCAF, MYST, p300/CBP and RTT109 (Marmorstein and Zhou, 2014). All of them catalyze the transfer of an acetyl group from acetyl-CoA to an acceptor lysine residue. Acetylated lysine is recognized by bromodomain containing proteins while removal of acetylation is mediated by histone deacetylases (HDAC). For most of the other histones modifications writers and eraser enzymes as well as readers have been described and their number is constantly growing. Because of correlations between several human diseases and histone PTMs balance, proteins that mediate histone PTMs as well as readers of histone PTMs have become attractive drug targets.

Another layer of complexity in chromatin organization is represented by the presence of several histone variants that differ in the amino acid sequence from the canonical histones and are linked to specific functions such as transcriptional activation (H3.3),

kinetochore assembly (CENPA), DNA repair and recombination (γH2AX) or X chromosome inactivation (macroH2A) (Sarma and Reinberg, 2005).

1.1.3 DNA methylation

Besides histones also DNA can be covalently modified. The most common modification in eukaryotes is the methylation of cytosine at the carbon at the 5th position of the pyrimidine ring. Since it does not affect the Watson-Crick base-pairing of cytosine, 5-methylcytosine (5meC) is a classical epigenetic modification and it is involved in different epigenetically regulated processes like imprinting, X chromosome inactivation and silencing of repetitive sequences. This modification takes place mainly at CpG dinucleotides and plays a role in the regulation of gene expression (Jones, 2012). DNA methylation at GC-rich promoters and close to transcription start sites (TSS) is often correlated with transcriptional repression while 5mC is commonly found in the gene body of transcriptionally active genes (Jones, 1999). Roughly 80% of the CpGs are methylated in mammal genomes and this modification is particularly enriched at repetitive elements, retrotransposons, gene bodies and satellite DNA (Ehrlich et al., 1982; Li and Zhang, 2014). Clusters of CpG called CpG island (CGI) are usually found overlapping with promoter sequences in particular of housekeeping genes and are mostly unmethylated (Deaton and Bird, 2011). Methylation of CGI is usually found at promoters of imprinted genes or, in somatic cells, at promoters of genes exclusively expressed in germ cells and in both cases it is linked to transcriptional repression (Borgel et al., 2010; Li et al., 1993). The acquisition of a genome wide program of CpG methylation is essential to determine the exit from the pluripotent state of embryonic stem cells and to allow differentiation (Gifford et al., 2013; Mohn and Schubeler, 2009; Shipony et al., 2014; Smith et al., 2012). Cytosine methylation is catalyzed by a group of enzymes called DNA methyltransferase (DNMT). In particular DNMT1 has been described as the enzyme responsible for maintenance of CpG methylation after replication (Bestor, 2000). Upon the passage of the replication fork DNMT1 re-establishes 5meC on the newly synthesized strand thus allowing faithful maintenance of the DNA methylation pattern over generations (Hermann et al., 2004). DNMT3a and DNMT3b carry out *de novo* establishment of CpG methylation acting preferentially on unmethylated DNA (Yokochi and Robertson, 2002). It has been shown that all the DNMTs are essential

for correct development both in mouse and in human. Mutations affecting human DNMT genes have been linked to several diseases like acute myeloid leukemia (AML) and immunodeficiency, centromere instability and facial abnormalities (ICF) syndrome (Jin et al., 2008; Shah and Licht, 2011). DNMT1 knockout is embryonic lethal in both species and even tough mouse *Dnmt1*-/- embryonic stem cells (ESCs) have been obtained these show global loss of DNA methylation and severe genomic instability (Guo et al., 2004; Kim et al., 2004). *Dnmt3a* and *Dnmt3b* double knockout (dKO) is embryonic lethal in mouse and single KO leads to embryonic (*Dnmt3b*) or postnatal lethality (*Dnmt3a*) (Okano et al., 1999). Surprisingly, in the same work *Dnmt3a* and *Dnmt3b* dKO ESCs have been obtained but they displayed absence of *de novo* methylation activity, suggesting that *de novo* methylation is actually needed later than blastocyst stage during embryonic development. Very recently *DNMT* KO human ES cells have been obtained by means of CRISPR/Cas9 genome editing technology (Liao et al., 2015). Remarkably *DNMT3a* and *b* single or dKO did not show defect in pluripotency being able to form teratomas when injected in nude mice, but, as expected, had a severe impairment of genome wide *de novo* methylation. In the same work it was shown that DNMT1 is essential for human ES cell viability since it was impossible to obtain a *DNMT1*-/- stable cell line and even conditional KO cells underwent massive cell death upon induction of recombination.

DNA methylation is a stable epigenetic modification but it can be also removed passively or actively. Passive DNA demethylation is achieved during successive rounds of DNA replication in which DNMT1 is inhibited or has reduced activity (Wu and Zhang, 2010). Active DNA demethylation is still a controversial topic since several enzymes and different pathways have been described to be involved in the mechanism (Ooi and Bestor, 2008). Nowadays the most acknowledged active mechanism involves the iterative oxidation of 5mC into 5-hydroxymethilcytosine (5hmC), 5-formylcytosine (5fC) and finally 5-carboxylcytosine (5caC) by the ten-eleven translocation (TET) enzymes. 5fC and 5caC can be eventually removed by the thymine DNA glycosylase (TDG, normally involved in T:G mismatch repair) in the context of the base excision repair (BER) pathway, allowing restoring the canonical C:G base pair (He et al., 2011; Kohli and Zhang, 2013). A good example in which active and passive demethylation mechanisms occur together is immediately after oocyte fertilization in mammals. At this stage both paternal and maternal genome are widely methylated but immediately undergo substantial demethylation. The maternal

DNA goes through passive demethylation during following cell divisions while the paternal genome is first massively hydroxymethylated by the maternally stored TET3 enzyme at the zygote stage and then the oxidation products are diluted through a replication dependent process (Inoue and Zhang, 2011).

As in the case of histone modifications, also "readers" of methylated DNA have been described. These proteins contain a methyl-CpG binding domain (MBD) and are involved in the recruitment of additional factors that mediate transcriptional repression (Fatemi and Wade, 2006). As an example the methyl-CpG binding protein 2 (MeCP2) is able to recruit both HDACs and the HMTs SUV39H1 and 2 (which catalyze H3K9me3) to establish histone repressive marks and transcriptional repression (Fuks et al., 2003; Nan et al., 1998). From the opposite point of view also recruitment of DNMTs by histone modifications or histone modifiers have been described. The PWWP domain of DNMT3B is for instance able to recognize H3K36me3, which is co-transcriptionally deposited by SETD2 at the gene body of active genes, and this can explain the recruitment of DNMT3B and the deposition of *de novo* DNA methylation at these sites (Baubec et al., 2015). The G9a HMT, which is able to mediate H3K9me1, me2 and me3, is also able to directly recruit DNMT3A and DNMT3B by an ankyrin domain independently of the HMT activity reinforcing the transcriptional repression of pluripotency related genes during ESC differentiation (Epsztejn-Litman et al., 2008). DNA methylation can also inhibit recruitment of DNA binding factors. For instance in mouse cells, methylation of a single CpG within the UCE (upstream control element) of the ribosomal gene promoter located 133 bp upstream the TSS impairs binding of the Pol I transcription factor UBF (upstream binding factor) to chromatin, thereby preventing the formation of the initiation complex (Santoro and Grummt, 2001). These represent only few examples of the interplay between DNA methylation and histone modifications showing how these epigenetic mechanisms are actually deeply interconnected to modulate gene expression.

1.2 Non-coding RNAs

Complexity of organisms does not directly correlate with the size of their genome (C-value paradox) neither with the number of their genes (G-value paradox) (Hahn and Wray, 2002; Schad et al., 2011). For example according to the animal genome

size database (http://www.genomesize.com) the human genome has the same size as the common frog (*Rana temporaria*) and contains roughly the same number of protein-coding genes of the plant *Arabidopsis thaliana* (Pertea and Salzberg, 2010). Even taking in consideration events such as alternative splicing of transcripts or post-translational modifications of proteins it is not possible to justify the differences in organismal complexity among these species. Thus it has been proposed that the complexity of an organism better correlate with the degree of control of gene regulation. This is obtained, in higher organisms, by expansion of *cis*-acting regulatory elements that can regulate transcription but also splicing (Fu and Ares, 2014), chromatin architecture (Fennessy and Owen-Hughes, 2016; Yang et al., 2016), RNA editing (Sapiro et al., 2015) and RNA stability (Taft et al., 2007). A further level of complexity in higher organisms is reached by the increase of transcription of non-protein-coding DNA that can give rise to a wide range of regulatory RNAs, which directly exert their function as established molecular entities without being translated into protein. Indeed despite only a small portion (ca. 2%) of the mammalian genome is transcribed into messenger RNAs encoding for proteins, the vast majority (ca. 80%) is actually transcribed (Bertone et al., 2004; Carninci et al., 2005; Cheng et al., 2005; Djebali et al., 2012). With the only exception of ribosomal RNA and transfer RNAs, whose function was revealed already in the early 1970s (Rich and RajBhandary, 1976; Schweet and Heintz, 1966), this big amount of ncRNA has been considered in the past the result of "transcriptional noise" of what was considered at that time "junk" DNA and by analogy considered non-functional "junk" RNAs. Nowadays several ncRNAs have been described to have a biological function that ranges from regulation of transcription, mRNA degradation, repression of translation, regulation of mRNA splicing, scaffold for ribonucleoprotein (RNP) complex formation, chromatin regulation and nuclear architecture. This wide range of functions as well as the genomic origin has challenged any effort of classifying ncRNAs into functional or structural groups. A rudimentary classification relies on their size distinguishing between short (or small) and long ncRNA (lncRNA) if respectively shorter or longer than 200 nucleotides (Clerget et al., 2015; Kung et al., 2013).

1.2.1 Small non coding RNA

tRNAs, snRNAs snoRNAs, miRNAs, siRNAs, piRNAs and tiRNAs belong to the small ncRNAs class (Clerget et al., 2015). The well-characterized transfer RNAs (tRNAs) represent the cornerstone of the mechanism of translation of the genetic code into polypeptides chains (Schweet and Heintz, 1966). Small nuclear RNAs (snRNA) are found in the nucleoplasm where they exert their function. To this class belong the U1, U2, U4, U5 and U6 snRNA that are responsible for the formation of the spliceosome and the catalysis of intron removal from pre-mRNA (Valadkhan, 2005). Small nucleolar RNAs (snoRNA) were originally described in mediating modification and processing of ribosomal RNA inside the nucleolus. Indeed the two families of snoRNAs (C/D and H/ACA) guide respectively the 2'-*O*-ribose methylation and pseudouridylation of rRNA, snRNA and mRNA (Matera et al., 2007). Nevertheless snoRNAs, in conjunction with specialized proteins, are also involved in other processes outside of the nucleolus. As an example the telomerase RNA TERC belongs to the H/ACA snoRNA family and in combination with the reverse transcriptase (TERT) is involved in the synthesis and maintenance of telomeres, the structure that protects the terminal part of chromosomes (Collins, 2006). Small interfering RNA (siRNA; ca. 21 nt in length) and micro RNAs (miRNA ca. 22 nt in length) partially share the maturation and assembly machinery and are involved respectively in the active degradation and the inhibition of translation of the target mRNA (Bartel, 2004; Fire et al., 1998; Hutvagner et al., 2001). The main differences between these two classes of small ncRNA reside in the sequence complementarity with the target RNA that is perfect in the case of siRNA, while includes some mismatches in the case of miRNA, a second difference resides in the origin of the precursor double stranded RNA (dsRNA). miRNAs originate from miRNA genes whose transcript is partially self complementary and able to form a hairpin structure. This can be recognized and processed by the complex Drosha/DGCR8 into a pre-miRNA that is exported in the cytoplasm and further processed by Dicer to give rise to the mature miRNA, which in turn is loaded onto the RNA-induced silencing complex (RISC) including the Argonaute 1-4 proteins. siRNAs can originate from endogenous sources that are able to form dsRNA as for instance sense-antisense transcripts, self complementary RNAs, gene-pseudogene transcript pairs, or from exogenous sources like retroviral genomes. These transcripts are directly processed by Dicer in the cytoplasm of the cell and loaded onto the effector complex RISC, which

mediate the cleavage of the target RNA (Kim et al., 2009). Due to the high efficiency and specificity of target degradation siRNAs are nowadays a common tool to down-regulate protein expression in different experimental approaches. PIWI-interacting RNAs (piRNA; 24-31 nt in length) originate from intergenic repetitive elements mostly including transposable elements (retrotransposons), which are transcribed in single stranded RNA and processed by a not fully characterized complex that includes PIWI proteins. Mature piRNAs are then used in a ping-pong reaction that allows their amplification using the target retrotransposon RNA (Siomi et al., 2011). The main function of piRNA is indeed to silence transposable elements in the germ line both by degradation of the RNA product of retrotransposons and by inducing heterochromatin formation at genomic loci containing transposable elements (Iwasaki et al., 2016). Transcription initiation RNAs (tiRNA) are small ncRNA generated by stalling or backtracking of the RNA polymerase II near the transcription start site (TSS). Their function, if any, is still unknown and they have been proposed to be involved in maintaining chromatin into a transcriptionally active state (Aalto and Pasquinelli, 2012).

1.2.2 Long non coding RNA

In contrast to small ncRNAs, the heterogeneous and multiple functions of lncRNA still do not allow a simple classification into functional categories. Additionally, the study of lncRNA has revealed to be experimentally very challenging. In fact the genetic and biochemical approaches used for studying protein functions, which have been improved in several decades of experimental work, are not fully transferable to the study of lncRNA. For this reasons the existing techniques have to be wisely adapted to this scope and more likely new methods have to be established (Bassett et al., 2014; Leone and Santoro, 2016).

LncRNA have been implicated in several biological mechanisms like gene transcriptional regulation, chromatin modification, scaffold for RNP complexes formation, nuclear architecture and many other functions, which currently continue to be identified (Guttman and Rinn, 2012; Quinodoz and Guttman, 2014). A relatively easy way for classifying lncRNA relies on the distinction of their genomic origin (Kung et al., 2013). In particular, lncRNA can (i) originate from specific transcription units as in the case of long intergenic ncRNA (lincRNA); (ii) be the product of

antisense transcription of a protein coding gene or part of it; (iii) be transcribed from non-functional genes (pseudogenes); (iv) derive from introns of coding mRNA after splicing; (v) originate from sense or antisense transcription of promoter or enhancer regions. Even though quite realistic this classification probably does not cover all the possible sources of lncRNA and these categories are not mutually exclusive. Furthermore this classification does not take in consideration post-transcriptional modifications of RNA, like RNA editing or processing that could produce different variants of lncRNA or even new lncRNA derived for instance from processing of precursors with different functions (Savic et al., 2014).

As recently stated by Quinn and Chang, "perhaps the sooner we dispense with categorical definitions of lncRNAs and recognize that they exist on multidimensional spectra of biogenesis, form and function, the sooner we can appreciate the enormous diversity of these genes." (Quinn and Chang, 2016).

1.2.2.1 Mechanisms of transcriptional regulation mediated by lncRNA

LncRNAs can regulate transcription of protein coding genes acting both in *cis*, that is in proximity to the site of transcription of the lncRNA on the same allele or in *trans*, that is on different allele (Guttman and Rinn, 2012; Ma et al., 2013). *Cis*-acting regulation can be achieved as a consequence of the transcription of the lncRNA itself and result in activation or inhibition of the neighboring protein coding genes. *Cis*-transcriptional activation is thought to rely on the active chromatin state concurrent to the act of transcription of the adjacent or overlapping lncRNA gene and on the recruitment of transcriptional activators and chromatin modifiers that deposit active histone marks (Guil and Esteller, 2012; Krishnan and Mishra, 2014). Similarly transcriptional repression is hypothesized to result from collision of the machineries involved in the transcription of coding and non-coding genes respectively (Crampton et al., 2006), or from recruitment of transcriptional repressors and repressive histone marks writers (Guil and Esteller, 2012). *Trans*-acting regulation requires the diffusion of lncRNA from its site of transcription to the site of action in analogy to *trans*-acting protein factors. However lncRNAs, which are still tethered to their locus of origin, can also act in *trans* if the two genomic loci, the one of origin and the one of action, are in close proximity to each other (Guttman and Rinn, 2012). Therefore also the *cis*-*trans* classification is blurry due to the singular nature of lncRNAs. In a very naive

mechanistic classification lncRNAs can be described as signaling molecules, decoys, scaffolds and guides (**Figure 3**).

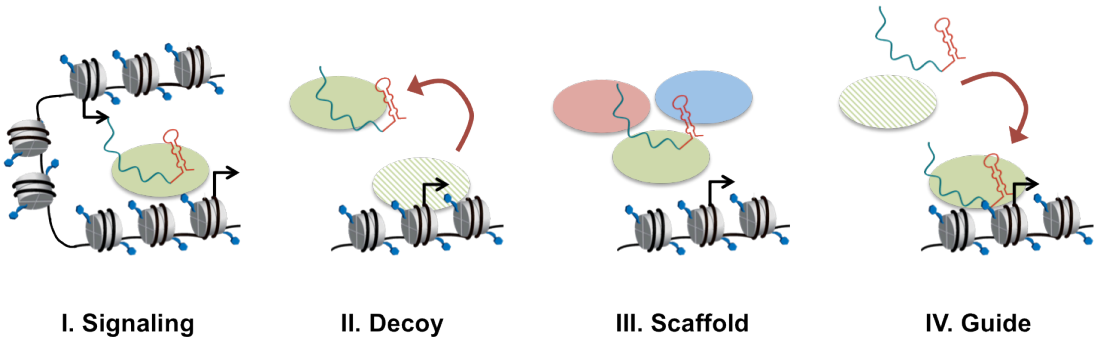


Figure 3 Schematic representations of the four archetypes of lncRNA mechanisms. Archetype I: as signals, lncRNA expression can reflect the combinatorial actions of transcription factors (colored ovals) or signaling pathways to regulate gene expression in space and time. Here is represented the chromatin loop formation mediated by an enhancer RNA. Archetype II: as decoys, lncRNAs can titrate transcription factors and other proteins away from chromatin or titrate the protein factors into nuclear subdomains. A further example of decoys is lncRNA decoy for miRNA target sites (see section 1.2.2.4). Archetype III: as scaffolds, lncRNAs can bring together multiple proteins to form ribonucleoprotein complexes. The lncRNA-RNP may act on chromatin to affect histone modifications. In other instances, the lncRNA scaffold is structural and stabilizes nuclear structures or signaling complexes. Archetype IV: as guides, lncRNAs can recruit chromatin-modifying enzymes to target genes, either in *cis* (near the site of lncRNA production) or in *trans* to distant target genes. All the archetypes can result in activation or repression of gene transcription according to the function of the associated proteins. Adapted from (Wang and Chang, 2011)

LncRNA can act as signaling molecules being transcribed under certain external stimuli and contributing in gene expression. An example of signaling lncRNAs is enhancer RNAs, which can regulate transcription determining chromatin looping. LncRNAs can act as decoys being able to bind and sequester protein targets that are involved in transcriptional regulation. LncRNAs can function as molecular scaffolds to build RNP complexes and can act as guide if they localize these RNP complexes or single factors to specific genomic sites (Wang and Chang, 2011). Also all these functions are not mutually exclusive and a single lncRNA can possess all these functions at the same time. In the recent years many models for lncRNA function have been proposed. I will here describe some well-studied examples of lncRNA-mediated regulation, which highlight the broad and multifaceted features of lncRNA mechanisms of action.

1.2.2.3 LncRNAs in the regulation of sex chromosome dosage compensation: *Xist* and *roX* RNAs

An important example of lncRNAs functions is represented by the mechanism through which many animals compensate the aneuploidy of one of the sex

chromosomes in order to maintain a balanced expression of the genomes of the two sexes, defined as sex chromosome dosage compensation (DC). In mammals and fruit flies two X chromosomes characterize females, while males have one X and one Y chromosome (Brockdorff and Turner, 2015; Lucchesi and Kuroda, 2015). In female mammals, DC is achieved by the transcriptional inactivation of one of the two X chromosomes at early stages of embryonic differentiation (X chromosome inactivation, XCI). In fruit flies the opposite mechanism takes place that means over expression of the single X chromosome in males. Both mechanisms are mediated by lncRNAs: Xist in mammals and roX1 and roX2 in *Drosophila.*

In mammals the X inactive specific transcript (Xist) is transcribed from the X inactivation center (Xic) of the X chromosome that will be inactivated and is able to recruit *in cis* chromatin-modifying complexes that are responsible for silencing, formation of heterochromatin and compaction of the X chromosome. In mice Xist is a 15 Kb lncRNA transcribed by RNA pol II, spliced, capped and polyadenylated (Borsani et al., 1991; Brockdorff et al., 1992). Its sequence is poorly conserved even among mammals but its structure is highly similar in different species and is characterized by tandem repeats dubbed motif A to F that determine its ability to act as a scaffold and as a guide (Brockdorff et al., 1992; Brown et al., 1992). Indeed the C motif is essential for the correct targeting of Xist, while the A motif is necessary for induction of silencing through binding of specific factors (Sarma et al., 2010; Wutz et al., 2002). The mechanism by which Xist determines silencing of the X chromosome is still under debate. RNA immunoprecipitation (RIP) experiments showed that the PRC2 subunits EZH2 and SUZ12 co-immunoprecipitate with Xist RNA, suggesting that Xist can act as a scaffold for the recruitment of the PRC2 repressive complex on the X, which results in the observed accumulation of H3K27me3 across the inactive X (Zhao et al., 2010; Zhao et al., 2008). However recent studies using more stringent techniques to purify RNA-protein complexes, which include a UV-crosslinking step, identified others factors that interact more tightly with Xist and can mediate XCI. Among those factors HNRNPK was shown to bind the F motif in exon 1 of Xist and mediate the recruitment of PRC2 complex (Chu et al., 2015). Additionally SHARP (SMRT and HDAC associated repressor protein, also known as SPEN) was found to directly bind the A repeat, linking Xist targeting to histone deacetylation activity of HDAC3, which is a fundamental step of Xi (Chu et al., 2015; McHugh et al., 2015).

At the onset of XCI, Xist is transcribed from the Xic of the future Xi and rapidly spreads and coats the entire chromosome through a not well-known mechanism. Bivalent proteins like YY1 and HNRNPU (also known as SAF-A), which are able to bind both RNA and DNA, have been described to mediate Xist localization on the Xi (Hasegawa et al., 2010; Jeon and Lee, 2011). However spreading of Xist across the X chromosome has been recently described to rely on spatial proximity of the Xic to some regions of the X chromosome rather than on sequence specificity (Engreitz et al., 2013). According to this model Xist exploits the pre-existing three-dimensional conformation of the X chromosome to identify target sites at distal regions, binding preferentially at the periphery of gene rich regions. This allows to recruit PRC2, HDAC and other factors responsible for silencing and compaction of chromatin and its repositioning in the Xist RNA compartment close to the nuclear lamina (Chen et al., 2016). In this way new chromosome regions are brought in close proximity to the Xic allowing the spreading of heterochromatin across the entire X chromosome.

Dosage compensation in Drosophila implies up regulation of the single X chromosome in males to balance the sex chromosome/autosome ratio. In this case DC is mediated by the complex male specific lethal (MSL) composed by the histone acetyltransferase MOF (males absent on the first), the scaffold protein MSL1, the E3 ubiquitin ligase MSL2, the chromodomain containing protein MSL3, the RNA/DNA helicase MLE (ortholog of human RHA also known as DHX9) and the two lncRNAs RNA on X 1 and 2 (roX1 and roX2) (Lucchesi and Kuroda, 2015). In particular roX1 and roX2 have a fundamental role in DC. Both are transcribed from the X chromosome and differ greatly in size (respectively 3.7 and 0.6 Kb) and sequence but share a common secondary stem loop structure called roX-box that could explain their redundant function (Meller and Rattner, 2002). This secondary structure is actively remodeled by the RNA helicase MLE to allow its recognition by the MSL2 protein during MSL complex recruitment on the X chromosome (Ilik et al., 2013; Maenner et al., 2013). The mode by which the MSL complex is recruited on the X chromosome is not fully understood and it seems to depend both on the ATP-dependent remodeling of roXs stem-loops by MLE and on the direct binding of the complex roX-MLE-MSL2 to high affinity DNA sites (HASs) acting as a platform for recruitment of the other components of the MSL complex (Straub et al., 2013).

Both Xist and roX RNAs represent a perfect paradigm of *cis*-acting non-coding transcripts that can function as scaffolds for the assembly of large RNP complexes

and that can guide these complexes to specific chromatin sites. Furthermore they act as signaling molecules since their expression is tightly controlled during development and involves complex regulatory networks that result in the modification of the chromatin status of entire chromosomes and in the establishment of a precise transcriptional program.

1.2.2.4 Further examples of lncRNAs functions

Though explaining a large number of lncRNA functions and mechanisms of action, the examples described so far cover only part of the broad and multifaceted features of lncRNA. Without the claim to provide a complete overview, a few other examples will be described.

HOTAIR, Air, Kcnq1ot1 and SRA are lncRNA that have been described to act as scaffold and guide for RNP complex. HOX antisense intergenic RNA (HOTAIR) originates from the HOXC gene cluster and acts *in trans* to repress the transcription of the HOXD genes. The mechanism of repression involves the interaction of the 5' region of HOTAIR with the PRC2 repressor complex and the 3' region with LSD1/CoREST complex in order to respectively establish the H3K27me3 repressive histone mark and remove the H3K4me3 active histone mark over the HOXD locus (Rinn et al., 2007; Tsai et al., 2010). Therefore HOTAIR is another important example of lncRNA acting as a modular scaffold and a guide for RNP complex formation. Air and Kcnq1ot1 are lncRNAs involved in regulation of imprinted gene repression. Both the RNAs are transcribed from the silenced paternal allele, and they specifically bind to and recruit *in cis* the histone H3 lysine 9 methyltransferase G9a to mediate H3K9me3 and transcriptional silencing of respectively *Kcnq1* or *Igf2r* loci (Nagano et al., 2008; Pandey et al., 2008). Furthermore the 5' end of Kcnq1ot1 RNA is responsible for binding and recruitment of DNMT1 to maintain CpG methylation of somatic differentially methylated regions (DMRs) gained during post-implantation development, while G9a and PRC2 recruitment by this RNA is necessary to mediate the imprinting of placental-specific imprinted genes (Mohammad et al., 2010). The lncRNA steroid receptor RNA activator (SRA) is another prominent example of modular scaffold RNA. SRA is indeed able to bind both PRC2 and Trithorax group (TrxG) complexes and in combination with the DEAD box RNA helicase p68 (also known as DDX5) and the MyoD transcription factor is involved in promoting muscle gene expression and cellular differentiation (Caretti et al., 2006; Wongtrakoongate et

al., 2015). Interestingly the p68/SRA complex is also involved in the stabilization of the interaction between cohesin and CTCF, promoting the insulator function of CTCF (Yao et al., 2010). Altogether these findings suggest that SRA may function as a scaffold to organize multiple factors, which in turn regulate gene expression in a context-specific manner.

Another function ascribed to some lncRNA is acting as decoy and sequestrate proteins in non-functional complexes. This is the case of growth arrest-specific 5 (GAS5), a lncRNA transcribed upon growth factor starvation. GAS5 contains a hairpin sequence motif that resembles the DNA sequence of the glucocorticoid response element (GRE) and this allows GAS5 to compete with GRE for the binding of the glucocorticoid receptor (GR). Thus upon starvation GAS5 is induced and act as a decoy for GR inhibiting the expression of glucocorticoid responsive genes (Kino et al., 2010). P21 associated ncRNA DNA damage activated (PANDA) is another lncRNA acting as decoy. PANDA is expressed in a p53 dependent manner and it binds to NF-YA transcription factor limiting expression of pro-apoptotic genes (Hung et al., 2011).

LncRNAs can also interact with other RNAs to exert their function. Two examples are represented by lincRNA-p21 and by TINCR. lincRNA-p21 was first discovered as a p53-induced lincRNA that acts as a transcriptional repressor by binding the nuclear factor KNRNPK, thus acting as a scaffold/guide to mediate gene silencing (Huarte et al., 2010). Subsequently lincRNA-p21 was also shown to form direct RNA-RNA interaction with beta catenin and JunB mRNAs in the absence of the RNA binding protein ELAVL1 and this resulted in inhibition of translation of these mRNAs mediated also by the interaction with the translational repressor Rck (Yoon et al., 2012). Another good example of post-transcriptional regulation mediated by lncRNA is the terminal differentiation-induced control RNA (TINCR). This cytoplasmic RNA controls human epidermal differentiation by stabilizing several mRNAs involved in skin development. Stabilization of those mRNAs requires both a direct RNA-RNA interaction, which involves a 25 nt motif called TINCR-box that is enriched in all the targets of TINCR, and the interaction with the RNA binding protein staufen1 (STAU1) (Kretz et al., 2013). Some lncRNA can act as sponges of miRNAs or proteins. To the first category belongs linc-RoR a competing endogenous RNA (ceRNA) specifically expressed in human ESCs. This RNA is transcribed under the direct control of the pluripotency factors Oct4, Sox2 and Nanog and it acts as a

competitor RNA for binding of the miRNA 145 (miR-145), which targets and down-regulate the transcripts of Oct4, Sox2 and Nanog via RISC mediated RNA degradation. Linc-RoR indeed contains in its sequence some miRNA-responsive elements (MRE) analogues to those of the pluripotency factors mRNAs. Thus miR-145 can target the MREs of linc RoR and this prevents degradation of Oct4, Sox2 and Nanog mRNAs and contributes to the maintenance of pluripotency (Wang et al., 2013). NORAD (ncRNA activated by DNA damage) represents a "protein-sponge" RNA. This RNA act as a signaling molecule that is massively induced upon genotoxic stress and is able to sequester PUMILIO proteins that specifically bind to RNA sequences called pumilio response elements (PRE), which are over-represented in NORAD sequence. In the absence of NORAD, PUMILIO proteins can bind to PREs present in the 3' untranslated region (UTR) of many mRNA encoding for factors involved in DNA repair and DNA replication. Binding of PUMILIO proteins to these targets determine their degradation and this results in genomic instability (Lee et al., 2016a).

The list of lncRNA described to have a biological function is continuously growing as well as the possible mechanisms through which they exert those functions. Certainly the field of lncRNAs is still in its infancy and much more has to be done and discovered to really appreciate the potential of this class of molecules. Nonetheless the evaluation of new possible mechanisms of action and biological functions of lncRNA has to be carefully and clearly proven in order to really understand the importance of these molecules. This consciousness is growing in the last years and most probably will lead to the revision of several mechanistic models that have been drawn so far using "old" biochemical and genetic approaches (Leone and Santoro, 2016).

LncRNAs have come a long way from being considered as "junk RNA" to their involvement in many biological processes (Cech and Steitz, 2014). Indeed they revolutionized the concept of transcriptome as a simple messenger of the genetic information towards the vision of it as an active entity that can take part in fundamental mechanisms of regulation of genome expression.

1.3 RNA helicases

Helicases are enzymes able to remodel nucleic acids by using the free energy of binding or hydrolyzing triphosphate nucleotides. All helicases, including both DNA and RNA helicases, are distributed among six superfamilies (SF1 to 6) according to their amino acid sequence and structure. All eukaryotic helicases belong to SF1 and SF2 while bacterial and viral helicases are distributed among all the SFs. RNA helicases are distributed in six families of which only one belongs to SF1 (Upf1-like) while the other five belong to SF2 (DExD, DExH/RHA, Ski2-like, RIG1-like and NS3/NPHII) (Fairman-Williams et al., 2010; Jankowsky, 2011) (**Figure 4**).

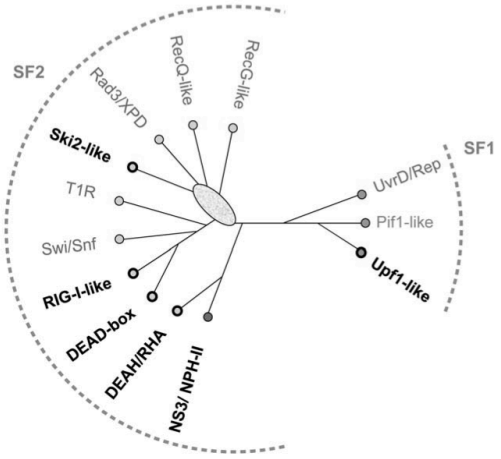


Figure 4. SF1 and SF2 helicase families. Unrooted cladogram showing the helicase families of the SF1 (right), and the SF2 (left). Branch lengths are not to scale. The oval indicates significant uncertainty in cladogram topology in this region. Boldfaced names show families harboring RNA helicases. Light grey names indicate DNA helicases. From (Jankowsky, 2011).

Helicases of SF1 and SF2 contain a structurally conserved helicase core region, formed by two highly similar RecA-like helicase domains arranged in tandem. Several RNA helicases of both SF1 and SF2 are actually able, at least *in vitro* to act both on DNA and RNA substrates. RNA helicase are believed to perform three different activities in the cell: they can unwind RNA duplex or more complex nucleic acid structures (proper helicase function); they can actively displace other proteins from RNA; and, in contradiction with their name, they can act in facilitating strand annealing and RNA folding (Jankowsky and Bowers, 2006; Jankowsky and Fairman, 2007). All these functions are thought to enable RNA helicases to catalyze disruption and formation of RNA secondary or tertiary structures and to remodel RNP complexes (Bhaskaran and Russell, 2007; Yang et al., 2007). Furthermore in some cases these reactions do not require ATP hydrolysis, which is instead more often necessary for efficient release of the helicase from RNA and thus for multiple substrate turnovers (Aregger and Klostermeier, 2009; Liu et al., 2008). Most, if not

all, RNA helicases work within multi-protein complexes and do not possess any sequence or structural preference. These aspects greatly complicate the identification of targets and more in general the study of their function *in vivo*. However sequence or site specificity would be most probably incompatible with the wide range of possible RNA folding or RNP complexes the RNA helicases have to deal with. Moreover there is a large lack of knowledge about the modulation of helicase enzymatic activity by other interacting proteins or by RNA structures.

The best studied RNA helicases belong to the DExD, DExH/RHA and Ski2-like families. Most of the studies on RNA helicases have been focused on mRNA splicing, mRNA translation, rRNA maturation and RNA degradation and very little is known about the relationship between RNA helicases and lncRNA (Bourgeois et al., 2016; Jarmoskaite and Russell, 2014). The DDX5/SRA complex already introduced in the section 1.4.2.4 represents one example of a functional RNA helicase/lncRNA complex. The interaction of DDX5 with the lncRNA SRA has been described to be important to coactivate MyoD-dependent transcription and, as a consequence, to enhance muscle differentiation (Caretti et al., 2006). DDX5/SRA is also important to mediate the interaction between the DNA-binding insulator protein CTCF and cohesin and to stabilize this interaction in order to achieve a proper insulator function (Yao et al., 2010). More recently DDX5 and SRA RNA have been shown to function as coactivators of Notch signaling by interacting with the CSL transcription factor and enhancing the expression of Notch target genes (Jung et al., 2013). DDX5 in conjunction with the lncRNA Rmrp also controls differentiation of T helper 17 lymphocytes (T_H17) by interacting with the RORγt transcription factor and coordinating the transcription of T_H17 specific genes (Huang et al., 2015). The RNA helicase DDX3x has been recently described in mice to inhibit maturation of specific miRNA in early postnatal retinal photoreceptors by inhibiting the processing of their pre-miRNA precursor. Expression of the lncRNA Rnrcr4 at later stages activate the processing of the pre-miRNA by directly interacting with the helicase DDX3x and allowing the maturation of the miRNA that regulates further steps of retinal development (Krol et al., 2015). The helicase DDX21 was described both to facilitate rRNA editing by interacting with snoRNAs and to interact with the 7SK snRNA, which sequesters the transcription elongator P-TEFb, determining P-TEFb release and enhancing Pol II transcription (Calo et al., 2015). All these works describe an intimate connection between RNA helicases and lncRNA in the modulation of several

processes, nevertheless they do not provide details of the mechanism through which helicases and lncRNAs can exert their biological function. As already mentioned above this can be ascribed to how little is known about the function of RNA helicases in the context of large multi-protein complexes and about their activity in modulating nucleic acid structures, which certainly goes beyond simple unwinding reactions.

1.3.1 DHX9 RNA helicase

The RNA helicase A (RHA, also known as DHX9 or NDHII) belongs to the SF2 and in particular to the DExH/RHA helicase family. As all the SF2 members DHX9 contain two very similar RecA-like domains, which are responsible for the helicase activity and the nucleotides triphosphate (NTP) hydrolysis and include the amino acid sequence DEIH (namely Asp-Glu-Ile-His) that is responsible for binding of NTP via Mg^{2+} (Lee and Pelletier, 2016). DHX9 is able to unwind *in vitro* both DNA and RNA by using the hydrolysis of any NTP (Zhang and Grosse, 1994). In addition to the helicase core domain, DHX9 contains two double stranded RNA-binding domains (dsRBDs) at its N-terminus. The minimal transactivation domain (MTAD), is situated between dsRBDII and motif I of the helicase core domain. A helicase-associated domain 2 (HA2) is adjacent to the C-terminal end of the helicase core domain. At the C-terminus are also present an oligonucleotide/oligosaccharide-binding fold (OB-fold), a nuclear localization signal (NLS) and a glycine rich RGG-box able to bind to single stranded nucleic acids with a higher affinity for ssDNA (Zhang and Grosse, 1997). DHX9 is well conserved among species and the human and murine homologues share 90% of amino acid identity while the Drosophila homologue MLE is 50% identical to the human DHX9.

DHX9 is a nuclear protein mainly localized in the nucleolus of mammal cells (Andersen et al., 2002; Fuchsova and Hozak, 2002; Zhang et al., 1999). Homozygous DHX9 deletion is embryonic lethal in mouse resulting in an aberrant gastrulation and massive cell death of ectodermal cells (Lee et al., 1998). However adult conditional knockout mice are not affected by the absence of DHX9 (Lee et al., 2016b). Deletion of the DHX9 homologue MLE in *Drosophila* it is lethal for male zygotes (Fukunaga et al., 1975). RHA knockout in *C.elegans* is linked to the loss of H3K9me3 in the germ line that leads to defects in mitosis and meiosis resulting in sterility (Walstrom et al., 2005). All these observations suggest that DHX9 is essential for development in

different organisms. Several years of research described DHX9 to be involved in different biological processes ranging from DNA replication, maintenance of genomic stability, mRNA splicing, miRNA biogenesis, RNA transcription and translation (Friedemann et al., 2005; Hartman et al., 2006; Hartmuth et al., 2002; Mischo et al., 2005; Nakajima et al., 1997; Robb and Rana, 2007). Furthermore DHX9 has been found implicated in several pathologies including HIV immunodeficiency, influenza A, systemic lupus erythematosus (SLE) and different types of cancer (Fujii et al., 2001; Lin et al., 2012; Mills et al., 2013; Wei et al., 2004). Like for other RNA helicases the study of DHX9 is complicated by the absence of sequence specific targets and even more by its ability to act both on RNA and DNA. These assumptions make difficult to find a precise mode of action for this helicase and suggest that its substrates can virtually include any RNA able to form folded structures and any complex interacting with those RNAs. Furthermore the action of DHX9 on these complexes could result in remodeling of the RNA-protein interaction or in its destruction by displacing the protein from the RNA molecule. Lastly DHX9 could also bridge the interaction between a RNA molecule and a protein being able to bind to both.

1.4 Embryonic stem cells (ESCs)

Embryonic stem cells (ESCs) derive from the inner cell mass (ICM) of a blastocyst, an early stage of mammalian development corresponding to the pre-implantation embryo (in mouse embryonic day 3.5, (E3.5)). ESCs are characterized by the ability to differentiate in any of the three germ layers that constitute the embryo of vertebrates (i.e. endoderm, mesoderm and ectoderm). This trait is defined as pluripotency. Stemness is considered the ability of a cell to self-renew through symmetric division, originating identical daughter cells, and to differentiate through asymmetrical divisions from which originate a daughter cell with stem features and a daughter cell that is “committed” to differentiate. The potency of a stem cell reflects the variety of cell types that can potentially originate by differentiation. The potency ranges from the totipotency of the zygote, from which originate both embryonic and extra-embryonic tissues, to the unipotency of some adult stem cells such as epithelial stem cells, which guarantee cell turnover in epithelia (Slack, 2000).

Pluripotency together with the self-renewal ability, which means the capability to proliferate and maintain an undifferentiated state, make ESCs a perfect tool to dissect the molecular mechanisms of cell differentiation and early development.

1.4.1 A short history of ESCs

ESCs were originally derived in the early '80s from the ICM of early mouse embryos at E3.5 and cultured on a "feeder" cell layer (mitotically inactivated fibroblast) in medium containing fetal bovine serum (Evans and Kaufman, 1981; Martin, 1981). The efficiency of those early attempts was quite low and further refinements of the culture conditions allowed establishing more stable and homogenous cell lines. A big improvement was achieved by the introduction of the leukemia inhibitory factor (LIF) in the culture medium (Smith et al., 1988; Williams et al., 1988). LIF was isolated from the cytokine factors produced by the feeder cells and it is a member of the IL6 family that binds to the leukemia inhibitory factor receptor (LIFR also known as CD118). This binding induces the heterodimerization of the LIFR with the glycoprotein 130 (gp130) triggering three different signaling pathways. In particular the JAK (Janus Kinase)/STAT3 (signal transducer and activator of transcription 3) pathway controls the transcription of genes regulating self-renewal; the PI3K (phosphatidylinositol-3-OH kinase)/PKB (protein kinase B) promotes ESC survival and cell cycle progression by suppressing the activity of the cell cycle inhibitor p27/kip1; the SHP2 (SH2 domain containing tyrosine phosphatase)/MAPK (mitogen activated protein kinase) pathway induces a phosphorylation cascade of extracellular-signal-related kinases (ERKs) which leads to activation of genes involved in differentiation. The SHP2/MAPK pathway counteracts the self-renewal signaling of the JAK/STAT pathway, thus inhibition of MAPK by the small molecule PD0325901 promotes self-renewal (Burdon et al., 1999). The use of this MAPK inhibitor in conjunction with the Wnt signaling activator CHIR99021 and the bone morphogenic factor 4 (BMP4), which allow omitting serum in the culture medium, represents one of the most common serum-free culture conditions used to maintain ESCs in culture, commonly referred to as 2i condition (Wray et al., 2010; Ying et al., 2008).

Pluripotent stem cells can be derived from other sources and using different techniques. Epiblast stem cells (EpiSCs) can be obtained for instance from post

implantation epiblast at E5.5-6.5. These cells are able to form teratomas if injected in nude mice but are not able to efficiently contribute to chimera embryos if injected into blastocysts (Hayashi et al., 2011). Embryonic germ (EG) cells were derived from primordial germ cells (PGCs) from early-somite stage embryos (E7.5-13.5) (Matsui et al., 1992). Another approach to obtain stem cells is to induce nuclear reprogramming, which consists in the erasure of the epigenetic state of terminally differentiated cells to induce the acquisition of a pluripotent state that can direct the development of a complete new organism. One technique that allows nuclear reprogramming is the somatic cell nuclear transfer (SCNT), which involves the microinjection of the nucleus of a donor somatic cell into an enucleated oocyte (Wilmut et al., 1997). Reprogramming of somatic cells to an ES-like state can be achieved also by overexpression of the so-called "Yamanaka factors" namely the transcription factors Oct4, Nanog, c-Myc and Klf4 (Takahashi and Yamanaka, 2006). This approach and others derived from it (Malik and Rao, 2013) allowed the generation of induced pluripotent stem cells (iPSC). This technology revolutionized the concept of personalized medicine giving the opportunity to produce patient specific iPSC that can be used to model personalized therapeutic approaches and for cell transplantation strategies (Chun et al., 2010; Oh et al., 2012; Sun et al., 2010).

1.4.2 Transcriptional network of pluripotency

Pluripotency is maintained through a complex network of transcription factors that activate self-renewal related genes and represses pathways that lead to differentiation. Some of the master regulators of this network have been identified in the last years and in particular Oct4, Sox2 and Nanog are considered the pluripotency core factors essential to maintain the undifferentiated state. The *Pou5f1* gene encodes for OCT4 that belongs to the Pic-Oct-Unc (POU) family of transcription factors. In mouse OCT4 is exclusively expressed in the totipotent blastomeres, the pluripotent epiblast and in the PGCs (Rosner et al., 1990; Scholer et al., 1990). SOX2 belongs to the high mobility group (HMG) DNA binding proteins and it is expressed within the ICM and the extraembryonic ectoderm of pre-implantation embryos (Avilion et al., 2003). Both OCT4 and SOX2 are essential to establish and maintain pluripotency since *Oct4* or *Sox2* KO embryos are not able to form a pluripotent ICM and differentiate into the extraembryonic tissue trophectoderm (Avilion et al., 2003; Masui et al., 2007).

NANOG is able to support ESC self-renewal in the absence of LIF/Stat3 pathway. Even though *Nanog* KO embryos do not possess a pluripotent ICM, *Nanog* KO ESC can be generated afterwards. They are still pluripotent though prone to differentiate and are still able to contribute to chimeras but cannot give rise to PGCs. This suggests that NANOG is essential to establish the pluripotent state of the ICM and to safeguard the pluripotent status of the germ line (Chambers et al., 2007; Mitsui et al., 2003). Several genome-wide approaches described the genomic distribution of the core pluripotency factors and revealed that these factors co-occupy the promoters of several other transcription factors regulating their expression. In particular OCT4, SOX2 and NANOG bind to genes involved either in maintenance of pluripotency or in establishing differentiation programs. The binding to these promoters exerts differential regulation and determines repression of transcription of genes involved in differentiation and maintenance of an active state of pluripotency related genes. Interestingly the core pluripotency factors also co-occupy their own promoters and enhance their own transcription thereby enforcing the maintenance of pluripotency and self-renewal (Boyer et al., 2005; Loh et al., 2006). Thus, OCT4, NANOG and SOX2 control a cascade of pathways that are intricately interconnected and control pluripotency, self-renewal and cell fate determination (Loh et al., 2006; Wang et al., 2006).

1.4.3 The epigenetic state of ESC genome

The ability to give rise to any cell type of the embryo require a high grade of genome plasticity that means the ability to quickly establish a transcriptional program allowing for the selection of a specific path of differentiation. The chromatin state of ESCs has attracted considerable attention due to its distinct features. Indeed, chromatin in ESCs is increasingly being recognized as an open structure compared with chromatin in somatic cells, implying that its overall structure is less condensed and that the ratio between euchromatin and heterochromatin is higher than in differentiating cells (Bhattacharya et al., 2009; Meshorer and Misteli, 2006). The euchromatic state of the genome in ESC is associated with a wide range of active histone marks like H3K4me3, H3K27ac and H4ac that contribute to keep the chromatin accessible (Azuara et al., 2006; Gaspar-Maia et al., 2011; Liang and Zhang, 2013). This open genome conformation is thought to be linked to the elevated levels

of chromatin remodeling factors observed in ESCs. General transcription factors are also overexpressed in ESC compared to differentiated cells and this, together with an open chromatin conformation, allow stochastic formation of pre-initiation complexes determining overall elevated global genome transcription (Efroni et al., 2008; Guenther et al., 2007). Intronic and intergenic regions are in particular upregulated in ESCs as a consequence of the open chromatin state implying that at least some of these transcripts may act as non-coding RNAs that regulate pluripotency (Fatica and Bozzoni, 2014; Judson et al., 2009; Lin et al., 2009). Despite the transcriptional permissive state of the ESC chromatin a certain grade of gene silencing must take place in order to maintain pluripotency and genome stability. Genes related to differentiation pathways must indeed be in a silent state and transcription of sequences with a high mutagenic potential like retrotransposons must be repressed also in ESCs. Developmental genes are silenced in ESCs by the repressive histone mark H3K27me deposited by the EZH2 methyltransferase of the PRC2 complex (Reik, 2007). Interestingly some of these genes are also marked at their promoter with the active histone marks H3K4me3 and for this reason have been defined “bivalent” genes (Voigt et al., 2013). These bivalent domains are considered to poise expression of developmental genes thus allowing well-timed activation while maintaining repression in the absence of differentiation signals. Despite the existence of bivalent promoters is controversial, it provides an elegant mechanism to explain the plasticity of the ESC genome (Schmitges et al., 2011; Voigt et al., 2012). In contrast with developmental genes, which need to be loosely repressed, transposable element, or at least most of them, must be completely and tightly repressed in order to avoid deleterious genomic recombination events and this is achieved by H3K9 methylation and also by CpG methylation (Bourc'his and Bestor, 2004; Schlesinger and Goff, 2015).

ESCs differentiation involves a massive rearrangement of the epigenome that leads to repression of pluripotency genes and activation of specific subsets of developmental genes according to the differentiation program that is activated (Meissner, 2010). Remarkably, upon differentiation, ESC chromatin undergoes structural remodeling toward a highly condensed heterochromatic and transcriptionally repressed form (Bhattacharya et al., 2009; Meshorer and Misteli, 2006). The first line of evidence of this process came from the visualization of chromatin in ESCs using electron microscopy: heterochromatin was prevalent in

differentiated cells but much less in undifferentiated ESCs (Park et al., 2004; Savic et al., 2014) (**Figure 5**).

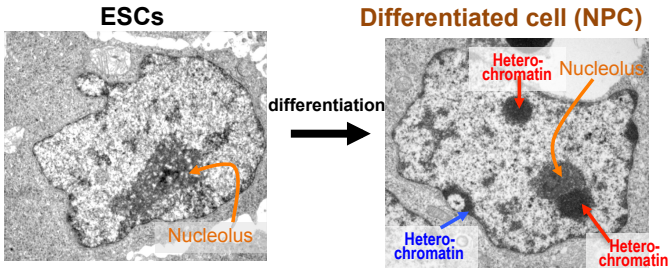


Figure 5. Reorganization of chromatin during ESC differentiation. EM images showing the formation of large heterochromatin blocks upon differentiation of ESCs into neural progenitors (NPCs). Arrows indicated the location of heterochromatin at nucleolus (red) and nuclear periphery (blue). Data from (Savic et al., 2014).

The majority of chromatin in ESCs is homogeneously spread and largely devoid of compact heterochromatin blocks, whereas in differentiated cells chromatin appears heterogeneous with distinct blocks of compaction (Efroni et al., 2008). Accordingly, while ESC chromatin fibers occupy the entire nuclear volume, the highly compacted chromatin of differentiated cells is organized into discrete domains leading to large regions of the nucleus devoid of DNA (Fussner et al., 2010). ESC differentiation is also accompanied by alterations of nuclear architecture such as formation of large organized chromatin regions enriched in the heterochromatic and repressive histone modification H3K9 methylation (termed LOCKs) (Wen et al., 2009), maturation and compaction of constitutive heterochromatin (such as centric and pericentric repeats) and clustering of highly condensed heterochromatin either at the nucleolus or at the nuclear periphery (Bartova et al., 2008; Wiblin et al., 2005) (**Figure 5**).

The open genome structure of ESCs well reflects the plasticity and transcriptionally permissiveness of ESC genome that has to have the ability to enter any distinct transcriptional programs for lineage specification (Gaspar-Maia et al., 2011; Gorkin et al., 2014) However, it still remains elusive how the switch from a lower to a higher order chromatin structure is achieved during ESC differentiation and whether this process plays a role in ESC differentiation.

1.5 Nucleolus and ribosomal genes

The nucleolus is the nuclear compartment where ribosomal RNA (rRNA) synthesis, rRNA maturation and ribosome biogenesis take place. Despite appearing as a dense structure in electron microscopy (EM) pictures, this membrane-less nuclear

subdomain is a highly dynamic structure that can disassemble and reform during each cell cycle around the nucleolus organizer regions (NOR) containing the rRNA gene clusters (Dundr et al., 2000; Shaw and Jordan, 1995). The nucleolus is composed of structural and functional subdomains that are visible in EM pictures. The fibrillar centers (FC) are the sites of active rRNA synthesis while the dense fibrillar components (DFC) contain newly synthesized pre-rRNA, which undergo modification, processing and partial assembly in pre-ribosomal particles. Later rRNA processing and maturation of ribosomes take place at the granular components (GC) (Scheer and Hock, 1999). Though the main function of nucleoli is to act as a "ribosome factory", several other tasks have been ascribed to this nuclear compartment (Olson et al., 2000). The nucleolus has been involved for instance in the assembly of ribonucleoprotein (RNP) machines like the signal recognition particle (SRP) (Politz et al., 2000), the spliceosomal small nuclear ribonucleoprotein (Kiss, 2001) and the telomerase (Etheridge et al., 2002). More recently it has been shown how the nucleolus and in particular the nucleolar heterochromatin plays a major role in shaping the nuclear architecture during ESC differentiation (Savic et al., 2014) and in anchoring the inactive X chromosome to perinuclear compartment in female mammals cells (Yang et al., 2015). Thus as stated by the Scottish embryologist C.H. Waddington already in 1965: "the nucleolus probably should not be considered a relatively simple organelle with a single function, comparable to a machine tool turning out a particular part of an automobile. It is not just 'the organelle where the cell manufactures ribosomes.' It is rather a structure through which materials of several different kinds are flowing, comparable more to a whole production line than to a single machine tool."

1.5.1 rRNA genes structure

Eukaryotic genomes contain several copies of rRNA genes, ranging from few hundreds in lower eukaryotes like *S.cerevisiae* to few thousands in some plants (Long and Dawid, 1980). These genes are organized in tandem arrays that are distributed among different chromosomes and each array can act as an independent NOR (Bell et al., 1992). Mouse rDNA is spread among the centromeric regions of chromosomes 12, 15, 16, 18 and 19 while in human is scattered among the short arm and the satellite body of the acrocentric chromosomes 13, 14, 15, 21 and 22 (Dev et al., 1977;

Henderson et al., 1972). Both mouse and human genomes contain roughly 400 copies of rRNA genes and each rDNA unit covers approximately 43 kb in human and 45 kb in mouse (Gonzalez and Sylvester, 1995; Grozdanov et al., 2003). Every array is composed of several dozens of precursor ribosomal RNA (pre-rRNA) genes separated by intergenic spacer sequences (IGS) (**Figure 6**). Transcription of a gene unit give rise to a pre-rRNA (45S in mouse and 47S in human), which is then modified and processed into 28S, 18S, and 5.8S rRNA. These rRNAs are then packaged with ribosomal proteins to form the scaffold and the catalytic heart of the large and the small subunits of ribosomes (Santoro, 2005).

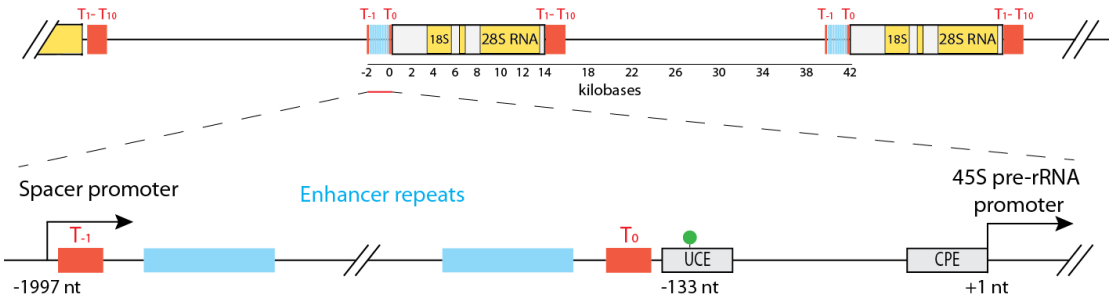


Figure 6. Organization of mammalian rDNA repeats. Black arrows indicate the sites of transcription initiation of the 45S pre-rRNA (TSS, transcription start site) and of the transcripts originating from the intergenic spacer promoter. Red boxes indicate terminator elements (T). Cyan boxes represent repetitive enhancer elements located between the spacer promoter and TSS proximal core promoter element (CPE). The upstream control element (UCE) located upstream the TSS is also shown. The green dot indicates the CpG at position -133, which is critical for rDNA silencing through DNA methylation.

Mouse rDNA promoter is composed of a core promoter element (CPE) neighboring the TSS and an upstream control element (UCE) roughly 100 bp upstream, which is recognized by the upstream binding factor (UBF) (**Figure 6**). UBF takes part in the assembly of the RNA Pol I preinitiation complex on the promoter together with the TBP (TATA binding protein) containing promoter selectivity factor TIFIB (SL1 in human) (Haltiner et al., 1986; Learned et al., 1986). Furthermore UBF is also involved in counteracting the repressive function of the heterochromatin protein 1 (HP1) at the rRNA genes, in stimulating the RNA pol I promoter escape and in promoting transcriptional elongation (Kuhn and Grummt, 1992; Panov et al., 2006; Stefanovsky et al., 2006). All these functions contribute in activating and promoting rRNA transcription. Methylation of the CpG at the position -133 inside the UCE impairs the binding of UBF to rDNA assembled into chromatin resulting in

transcriptional repression (Santoro and Grummt, 2001). An alternative rDNA promoter, dubbed spacer promoter, is present approximately 2Kb upstream the main rDNA promoter (Kuhn and Grummt, 1987). This element share some sequence homology with the main promoter and, despite having a reduced binding of TIFIB, it shows a threefold enrichment of Pol I compared to the main promoter (Santoro et al., 2010). From the spacer promoter originates an intergenic transcript that spans, codirectionally with the 45S pre-rRNA, through the intergenic sequence between the two promoters and overlap at its 3'end with the main promoter and part of the 5' sequence of the 45S pre-rRNA. This transcript, called intergenic spacer rRNA (IGS-rRNA), is readily processed, in differentiated cells, to give rise to a roughly 200 nt long non-coding RNA called promoter RNA (pRNA), which is involved in rRNA gene silencing (Mayer et al., 2006; Santoro et al., 2010) (see section 1.5.4).

Other *cis* regulatory elements called terminators (T) are present in ten copies (T_1-T_{10}) at the 3' end of each rRNA gene and are involved in transcription termination and in determining the directionality of the replication fork over the rDNA (Gerber et al., 1997; Grummt et al., 1986). Further T sequences are found upstream the rRNA gene promoter (T_0) and downstream the alternative spacer promoter (T_{-1}) and have been described to be involved in promoting rDNA transcription. T sequences are bound by the transcription termination factor 1 (TTF1), which was originally discovered as the factor responsible for the displacement of Pol I at the 3' end of rRNA genes and for termination of transcription (Grummt et al., 1986; Kuhn et al., 1990). However binding of TTF1 to the T_0 sequence was also proposed to enhance transcription through two distinct mechanisms. In one case T_0 bound TTF1 was shown to recruit an ATP dependent nucleosome remodeling activity close to the rDNA promoter that leads to the repositioning of the rDNA promoter associated nucleosome from an "off" to an "on" conformation upstream the TSS, which can allow transcription initiation (Langst et al., 1998; Langst et al., 1997). In the second case dimerization of TTF1 bound both at the T_0 and at the T_1-T_{10} sequences can determine the formation of a DNA loop that brings in contact the rDNA promoter and the 3' termination region facilitating in theory the direct transfer of Pol I from the termination site to the initiation site (Nemeth et al., 2008). The function of the T_{-1} sequence downstream the spacer promoter is still unknown. As it will be addressed in more detail in section 1.5.3, TTF1 bound to T_0 sequence is also able to recruit a repressor complex, which repositions the promoter nucleosome to the "off" location

and recruit several factors involved in silencing and formation of heterochromatin at a subset of rRNA genes.

Between the spacer promoter and the rDNA main promoter are located several repeated sequences that act as enhancer of rDNA transcription and are also bound by UBF (Pikaard et al., 1990). Different variants of rDNA units have been described according to the number of enhancer repeats (in mouse 6, 9, 10, 11, 12 and 22) and the specific subclass containing 9 repeats has been shown to be important for regulating the rRNA epigenetic status *in trans* (Santoro et al., 2010).

1.5.2 Chromatin organization of rRNA genes

In eukaryotic cells, a fraction of rRNA genes is transcriptionally silent. Electron microscopy visualization of rDNA units in *S. cerevisiae* (Miller spreads) revealed two different structures: (1) transcribing rRNA genes (active copies) that have a characteristic tree-like appearance (referred as "Christmas tree"), with a DNA "trunk" from which close-packed ribonucleoprotein "branches" of increasing length extend; (2) genes that do not associate with Pol I and are not transcribed (silent copies) (Miller and Beatty, 1969). Although the genome complexity of higher eukaryotes does not yet allow visualization of rDNA chromatin by Miller spreads, later biochemical studies assessed that the coexistence of active and silent rRNA genes in each cell is not limited to S. cerevisiae. Psoralen crosslinking, a technique that introduces crosslinks into DNA sites that are not protected by nucleosomes, showed the presence of a fraction of rRNA genes accessible to psoralen, thus largely devoid of nucleosomes and a second fraction that was not crosslinked by psoralen because densely compacted with nucleosomes. The demonstration that the nascent rRNA was associated with the psoralen crosslinked DNA led to the conclusion that the fraction of rDNA with reduced nucleosome content is actively transcribed while the nucleosome associated rDNA correspond to silent genes (Conconi et al., 1989; Sogo et al., 1984). Moreover, it was shown that the relative amounts of the two types of structures are similar in interphase and metaphase, indicating that active and silent rDNA repeats are maintained independently of the transcriptional process and are stably propagated through the cell cycle (Conconi et al., 1989).

Later studies showed that active and silent rRNA genes are also characterized by specific epigenetic marks. Silent genes are marked with CpG methylation at the UCE

and this leads to impairment of UBF binding (Santoro and Grummt, 2001; Stancheva et al., 1997). The finding that silent rRNA genes are enriched in CpG methylated sequences allowed for the distinction of silent and active rDNA chromatin and for the analysis of their chromatin composition. By using chromatin immunoprecipitation (ChIP) coupled with CpG methylation analysis employing methylation sensitive restriction enzymes (ChIP-chop) it was possible to identify protein factors and histones PTMs specifically associated with active (non CpG methylated) or silent (CpG methylated) genes (Santoro et al., 2002). This technique allowed showing that the promoter of mouse and human active rRNA genes is associated with Pol I transcription factors and histones modified with the active marks H4Ac and H3K4me2, while silent rRNA genes are associated with the heterochromatin protein 1 (HP1), TTF1-interacting protein 5 (TIP5), poly-ADP-ribose polymerase 1 (PARP1 also known as ARTD1) and with repressive histone marks H3K9me2, H3K27me3 and H4K20me3 (Guetg et al., 2012; Santoro and Grummt, 2001, 2005; Santoro et al., 2002) (**Figure 7**).

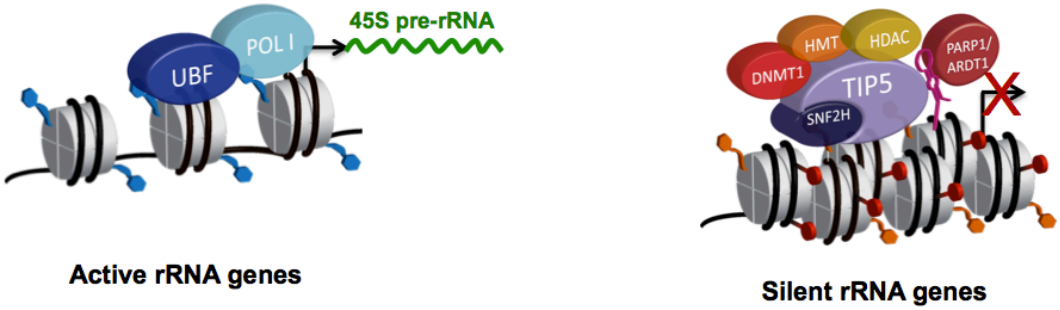


Figure 7. Schematic representation of active and silent rRNA genes. Transcriptionally active rRNA genes are characterized by open chromatin marked with active histone marks (blue lollipops) and are bound by UBF and Pol I. Transcriptionally silent genes are present in a compact heterochromatic structure marked with silencing histone marks (orange lollipops) and DNA CpG methylation (red lollipops). These repressive marks are ensured by the pRNA dependent binding of TIP5 to rDNA promoters and the consequent recruitment of histone modifiers and DNA methyltransferases (see sections 1.5.3 and 1.5.4).

1.5.3 The nucleolar remodeling complex (NoRC)

Previous studies have shown that the nucleolar remodeling complex (NoRC) is the key player in establishing and maintaining the heterochromatic state of silent rRNA genes. NoRC consists of TIP5 (TTF1-interacting protein 5) and the ATPase SNF2h (Li et al., 2005; Santoro et al., 2002; Strohner et al., 2001; Zhou et al., 2002). SNF2h is the human homolog of the Drosophila ISWI whose ATPase activity is essential for

nucleosome remodeling. TIP5 was identified in a yeast two-hybrid screening for TTF1 interacting proteins (Strohner et al., 2001). TIP5 is a member of the bromodomain adjacent zinc finger (BAZ) protein family and it is also known as BAZ2A (Jones et al., 2000). It shares several domains with other components of nucleosome remodeling complexes such as ACF, WCRF, CHRAC and WICH (Bochar et al., 2000; Bozhenok et al., 2002; Ito et al., 1999; LeRoy et al., 2000). Among these domains a bromodomain at the C-terminus allows the binding to acetylated histones and is necessary to mediate HDAC1 recruitment (Zhou et al., 2002), a plant homeodomain (PHD) allows the binding of SNF2h and other factors like HMTs and DNMTs (Zhou and Grummt, 2005; Zhou et al., 2002), a TAM (TIP5/ARBD/MBD) domain mediates the binding of RNA and in particular the stem loop structure of the lncRNA pRNA (Mayer et al., 2006). Other domains include a WAKZ motif, BAZ1 and BAZ2 motifs and several AT-hooks motifs. NoRC is targeted to the rDNA promoter via the interaction with pRNA and TTF1 (Mayer et al., 2008; Savic et al., 2014) and represses rRNA transcription by recruiting histone modifiers and DNA methylating activities (i.e., HDAC1, SETDB1, PARP1, DNMTs) (Santoro and Grummt, 2005; Santoro et al., 2002; Zhou et al., 2002; Zhou et al., 2009). The binding of NoRC to rRNA genes was shown to occur immediately after replication of silent rRNA genes in late S phase (Li et al., 2005), suggesting a role of NoRC in the maintenance and inheritance of silent rDNA chromatin during cell division. NoRC was shown to slide a nucleosome over the rDNA promoter in a sort of “inactive” position (from -157/-2 active to −132/+22 inactive) that inhibits rRNA transcription (Li et al., 2006). In this configuration the critical CpG dinucleotide located 133 bp upstream the TSS, whose methylation prevents binding of UBF, is placed at the 5’ boundary of the nucleosome and is exposed to methylation mediated by DNMTs associated with NoRC (Santoro and Grummt, 2001; Santoro et al., 2002). These findings suggest that NoRC can act through two distinct but not mutually exclusive mechanisms to silence rRNA genes: first by nucleosome remodeling at the rDNA promoter towards an “off” position and second by recruiting histone modifiers and DNA methylating enzymes to establish rDNA heterochromatin.

1.5.4 pRNA and the regulation of ribosomal genes

The lncRNA pRNA is critical for the establishment of silencing at rRNA genes. pRNA is a roughly 200 nt lncRNA that correspond to the rDNA main promoter sequence. pRNA is one of the few trans-acting lncRNA that have been described so far (Lee, 2012; Vance and Ponting, 2014; Yang et al., 2014). As it will be described in 1.5.5 pRNA originates from the processing of IGS-rRNA precursor, a transcript synthesized from the spacer promoter of a subset of active rRNA genes during early S phase (Mayer et al., 2006; Santoro et al., 2010). LNA-mediated knockdown of pRNA was shown to induce loss of heterochromatin at rRNA genes in differentiated cells, indicating the critical role of this lncRNA (Mayer et al., 2006). pRNA contains a secondary stem loop structure that is indispensable to bind the TAM domain of TIP5 and to recruit the NoRC complex to rDNA promoters (Mayer et al., 2006; Savic et al., 2014). Furthermore, the stem loop structure of pRNA acts as a scaffold to mediate the interaction of TIP5 with several factors including PARP1 and TTF1 (Guetg et al., 2012; Mayer et al., 2008; Savic et al., 2014). It was recently shown that mutations of pRNA sequences involved in the formation of stem loop structure, destroy TIP5-TTF1 interaction and abolish the recruitment of TIP5 to rDNA (Savic et al., 2014). Interestingly deletion of the 5' region of pRNA did not affect the interaction between TIP5 and TTF1 nor the recruitment of TIP5 to rRNA genes, indicating that the sole stem loop structure of pRNA is necessary and sufficient to allow TIP5 recruitment to rDNA and that the 5' sequence is actually dispensable (Savic et al., 2014). This result is in contrast with the model proposed previously for TIP5 recruitment to rDNA, which implicates the formation of a RNA-DNA triple helix through Hoogsteen or reverse Hoogsteen base pairing between the 5' sequence of pRNA and the corresponding T_0 sequence of the rDNA promoter (Bierhoff et al., 2014; Schmitz et al., 2010). Despite being attractive, this model has not been definitely experimentally proven and the finding that the stem loop structure of pRNA is the sole necessary sequence that TIP5 needs for its recruitment to rDNA depicts a more solid molecular model (Leone and Santoro, 2016).

1.5.5 IGS-rRNA transcription, pRNA maturation and maintenance of rRNA genes silencing

As already introduced in the section 1.5.1, RNA Pol I transcribes the 45S pre-rRNA starting from the rDNA main promoter and, in the same direction, the intergenic sequence rRNA (IGS-rRNA) starting from the spacer promoter that in mouse is located 2Kb upstream the rRNA gene main promoter. The IGS-rRNA sequence corresponds to the spacer sequence that separates spacer and main promoter and includes the sequence of the main promoter and part of the sequence downstream the TSS. The synthesis of IGS-rRNA has been shown to take place during a narrow time window of the early S phase specifically from a subset of active rRNA genes containing nine enhancer repeats between spacer and main promoter (Santoro et al., 2010). During mid-late S phase IGS-rRNA is processed into pRNA (Santoro et al., 2010). Interestingly the timing of IGS-rRNA transcription (early S phase) and its processing into pRNA (mid-late S phase) correlates with the observation that NoRC complex is recruited at mid-late S phase to newly replicated silent genes in order to maintain silent chromatin (Guetg et al., 2012; Li et al., 2005). Like other regions of the genome also rRNA genes experience a different replication timing according to the grade of chromatin compaction, being euchromatic genes replicated in early S phase and heterochromatic genes replicated in late S phase (Berger et al., 1997; Guetg et al., 2012; Li et al., 2005). Thus the fine-tuning of IGS-rRNA synthesis, pRNA maturation and NoRC recruitment represents an attractive model to explain how cells maintain the heterochromatic state of silent rRNA genes across cell divisions (**Figure 8**). Furthermore this model suggests that pRNA can act *in trans*, being transcribed from a subset of active rRNA genes and acting on late replicating silent rDNA to inherit DNA methylation and transcriptional repression.

The discovery that transcripts originating from the spacer promoter are necessary to silence rRNA genes is in apparent disagreement with previous studies showing that the spacer promoter enhances transcription from the main rDNA promoter (Caudy and Pikaard, 2002; Grimaldi and Di Nocera, 1988; Paalman et al., 1995). The enhancement of transcription has been explained with two mutually exclusive models. The first model is based on the concept of “read-through enhancement”.

In the “read-through enhancement” model, it was proposed that Pol I molecules, which are directed by the spacer promoter to transcribe through the enhancers, release rDNA transcription factors from the enhancers and make them available to the gene

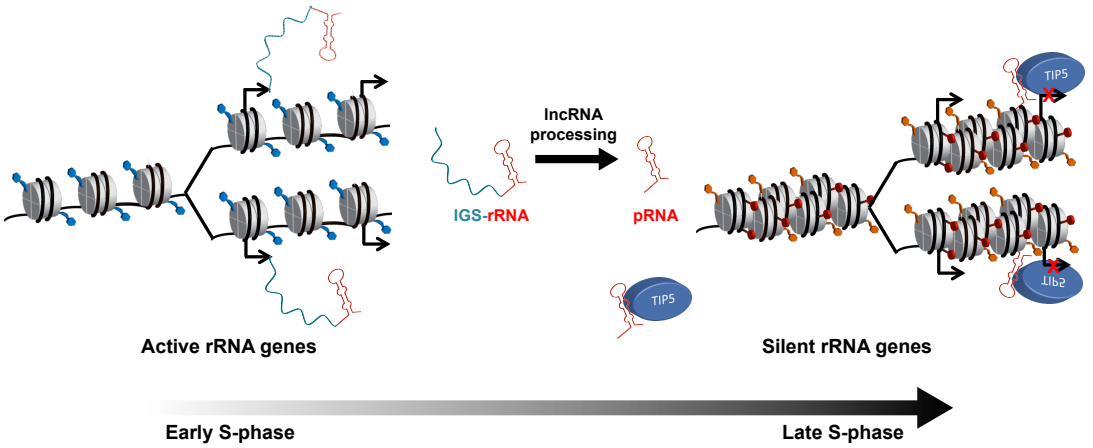


Figure 8. Model for the inheritance of silent rDNA heterochromatin. Transcriptionally active rRNA genes are replicated during the early S-phase. From these genes originates the IGS-rRNA that is processed during mid-late S phase giving rise to the mature pRNA. TIP5 associates with pRNA and is recruited to silent rRNA genes after their replication, which occurs at mid-late S-phase, ensuring the maintenance of the silent state by interacting with histone modifiers and DNA methyltransferase 1 (DNMT1).

promoter, thereby stimulating gene promoter transcription (De Winter and Moss, 1987). This model implies that transcription enhancement would be proportional to the strength of the spacer promoter. However, replacement of the mouse spacer promoter with the more active Chinese hamster spacer promoter did not increase rRNA transcription, indicating that enhancement of pre-rRNA synthesis does not depend on transcripts originating from spacer promoter and implying that spacer promoter affects the main gene promoter using alternative mechanisms (Paalman et al., 1995; Santoro, 2011). A second model proposed that it is the spatial proximity between spacer and main promoter through the formation of a DNA loop to be implicated in spacer-mediated enhancement of 45S pre-rRNA gene transcription (Santoro, 2011). The spatial juxtaposition of both promoters was proposed to enhance transcription from the main gene promoter by delivering Pol I factors without the requirement of IGS-rRNA synthesis. A similar model was proposed for the interaction between rDNA main promoter and terminator regions (Nemeth et al., 2008). As already mentioned in the section 1.5.1, TTF1 binds to T elements and is able to form dimers. Thus, two TTF1 molecules, bound respectively at T_{-1} and the T_0 element, might dimerize and bring spacer and main promoter in close proximity by forming a loop, in analogy to what has been described for rDNA promoter and terminator regions T_1-T_{10} (Nemeth et al., 2008). In this model, the involvement of TTF1 in forming the spacer-main gene promoter loop does not only suggest that IGS

rRNA synthesis is not required to enhance rRNA synthesis from the main promoter but also that it might not occur at all. Indeed, the ability of TTF1 to terminate Pol I transcription would represent a major obstacle for Pol I to elongate from spacer promoter toward the main gene promoter (Grummt et al., 1986; Kuhn et al., 1990). Consequently, in the "read-through enhancement" model TTF1 should not be bound to either T_0 and/or T_{-1} elements, an assumption that is in contrast with previous results showing that the binding of TTF1 to T_0 element is a prerequisite for 45S pre-rRNA synthesis (Langst et al., 1998). Taken together, all these observations suggest that the dual role of spacer promoter in regulating rRNA transcription can be distinguished by its capacity either to form a loop or to drive IGS rRNA synthesis: in the first case, it stimulates pre-rRNA synthesis; in the second case, it is required for NoRC-mediated rDNA silencing (Santoro, 2011).

1.5.6 Establishment of rDNA heterochromatin

Recent results have shown that, in contrast to somatic cells, in ESCs all the rRNA genes are active and devoid of heterochromatic marks (Savic et al., 2014). Though it is well known that ESC genome is highly euchromatic (see section 1.4) this observation was very surprising since it is also well known that reduction of rDNA heterochromatin in differentiated cells results in genomic instability due to recombination events between repetitive rDNA sequences (Guetg et al., 2010; Peng and Karpen, 2007; Straight et al., 1999). How stem cells can cope with such a high risk of deleterious recombination events is still not known.

It was recently shown that impairment of TIP5 association with rRNA genes is determinant for the lack of rDNA heterochromatin in ESCs. It is only upon differentiation that TIP5 re-localizes within nucleoli, binds and silences rRNA genes (Savic et al., 2014). This work also determined that a lncRNA-based mechanism mediates the switch of function of TIP5 (silencing rRNA genes in differentiated cells but not in ESCs). It was shown that in ESCs processing of IGS-rRNA into pRNA is impaired and it is only upon differentiation that this process is activated to produce mature pRNA. This study revealed that maturation of IGS-rRNA into pRNA is critical for the formation of rDNA heterochromatin. Indeed, transfection of mature pRNA into ESCs was sufficient to promote recruitment of TIP5 to rDNA and formation of rDNA heterochromatin. Importantly, Savic and colleagues showed that

only in its mature form pRNA allows the association of TIP5 with TTF1 at rDNA whereas the binding of TIP5 to the unprocessed transcript abolishes this process. Thus, the results described in Savic et al (Savic et al., 2014) indicated that the recruitment of TIP5 to rDNA promoter occurs via a *protein-RNA-protein-DNA* module (**Fig. 9**), which is in contrast with a previous model proposing that the T_0 sequence of pRNA form a triple helix with the T_0 element at rDNA promoter and is required to guide TIP5 to rRNA genes (Bierhoff et al., 2014; Schmitz et al., 2010). The analysis of pRNA T_0 sequence mutants did not reveal the involvement of triplex formation in TIP5 recruitment and rDNA silencing. In contrast, pRNA mutants with impaired stem-loop structure formation abolished the recruitment of TIP5 to rDNA. Taken together the results described in Savic et al. indicate that pRNA guides TIP5 to rDNA in trans through the hairpin structure that allows the interaction with TTF1. Thus, the impairment of IGS-rRNA processing, which abrogates the formation of mature pRNA, is the major determinant causing the euchromatic state of all rRNA genes in ESCs (Savic et al., 2014). How the activation of IGS-rRNA processing during ESC differentiation is regulated will be addressed in this thesis. Together these studies highlighted the role of lncRNA in the regulation of chromatin and epigenetic states and suggest that lncRNA processing represents an additional level of lncRNA regulation by modulating distinct features of the same lncRNA.

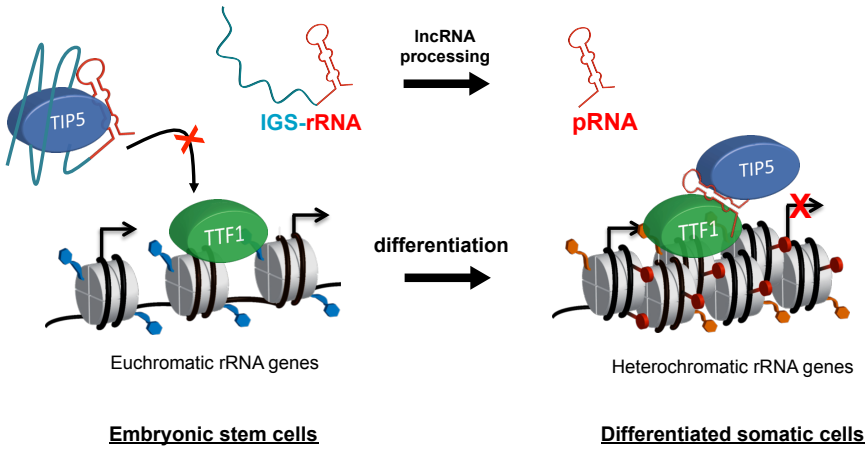


Figure 9. Model for the establishment of rDNA chromatin. In ESCs the IGS-rRNA is not processed and sequestrates TIP5 in an inactive complex. This results in the absence of rDNA heterochromatin. During differentiation IGS-rRNA is processed and the mature pRNA can accumulate, recruiting TIP5 to the rRNA genes promoter and establishing rDNA heterochromatin. Modified from Savic et al. 2014.

An important readout of Savic et al. was that the addition of mature pRNA in ESCs was not only sufficient to recruit TIP5 to nucleoli and silence rRNA genes but also induced the establishment of highly condensed chromatin structures outside of the nucleolus, resembling the genome organization that characterizes differentiated cells. These changes were also accompanied by a global increase in H3K9me2, maturation of heterochromatin at repetitive sequences - such as major and minor satellites - and their transcriptional repression as found in differentiated cells. Moreover, such heterochromatic ESCs were primed for differentiation due to up regulation of genes implicated in differentiation and developmental processes. Thus, these results suggest that the nucleolus is not only the cellular compartment where ribosomes are produced but it is also able to produce heterochromatin, affecting the genome architecture of the rest of the nucleus. The aim of this PhD thesis was to understand whether formation of heterochromatin in nucleolus plays a role for the ability of ESCs to enter into differentiation programs.

2 Aims

Increasing evidence indicates that genome architecture is important to ensure the correct execution of gene expression programs. An important example is provided by the spatiotemporal organization of ESC genome. The genome of ESCs is characterized by an overall open chromatin conformation, largely devoid of compact heterochromatin blocks, whereas in differentiated cells chromatin appears heterogeneous with distinct blocks of condensed structures. ESC differentiation is accompanied by alterations of nuclear architecture such as maturation and compaction of constitutive heterochromatin (i.e. centric and pericentric repeats) and clustering of highly condensed heterochromatin either at the nucleolus or at the nuclear periphery. The open genome structure of ESCs well reflects the plasticity and transcriptional permissiveness that that has to have the ability to enter any distinct transcriptional program for lineage specification. However, it still remains elusive how the switch from a lower to a higher order chromatin structure is achieved during ESC differentiation and whether this process plays a role in ESC differentiation.

Recent results from our laboratory (Savic et al., 2014) have shown that in embryonic stem cells rRNA genes are all euchromatic and acquire heterochromatic features only upon differentiation. Formation of rDNA heterochromatin during ESC differentiation is regulated by processing of the lncRNA IGS-rRNA into the mature pRNA and consequent recruitment of TIP5 to rRNA genes. Importantly, addition of mature pRNA in ESCs was not only sufficient to recruit TIP5 to nucleoli and silence rRNA genes but also induced the establishment of highly condensed chromatin structures outside of the nucleolus, resembling the genome organization that characterizes differentiated cells.

The aims of this work were to (1) identify factors implicated in IGS-rRNA processing and (2) determine the functional significance for the formation of rDNA heterochromatin during ESC differentiation.

3. Results

3.1 Research articles

3.1.1 Establishment of heterochromatin at rRNA genes is required for embryonic stem cell differentiation

Authors:	**Sergio Leone**, Dominik Bär, Coenraad Frederik Slabber & Raffaella Santoro
Journal:	Submitted for publication
Contribution:	Experimental design, performance and analysis of figures 1, 2, 4, 5, 6 and supplementary figures 1 and 3. S.L. wrote the manuscript together with R.S.

3.1.2 lncRNA maturation to initiate heterochromatin formation in the nucleolus is required for exit from pluripotency in ESCs

Authors:	Nataša Savić, Dominik Bär, **Sergio Leone**, Sandra C. Frommel, Fabienne A. Weber, Eva Vollenweider, Elena Ferrari, Urs Ziegler, Andreas Kaech, Olga Shakhova, Paolo Cinelli & Raffaella Santoro
Journal:	Cell Stem Cell
DOI:	10.1016/j.stem.2014.10.005
External link:	http://www.cell.com/cell-stem-cell/abstract/S1934-5909(14)00456-1
Contribution:	S.L. contributed to figures 2C, 4D, 5C

3.2 Review articles

3.2.1 Challenges in the analysis of long noncoding RNA functionality

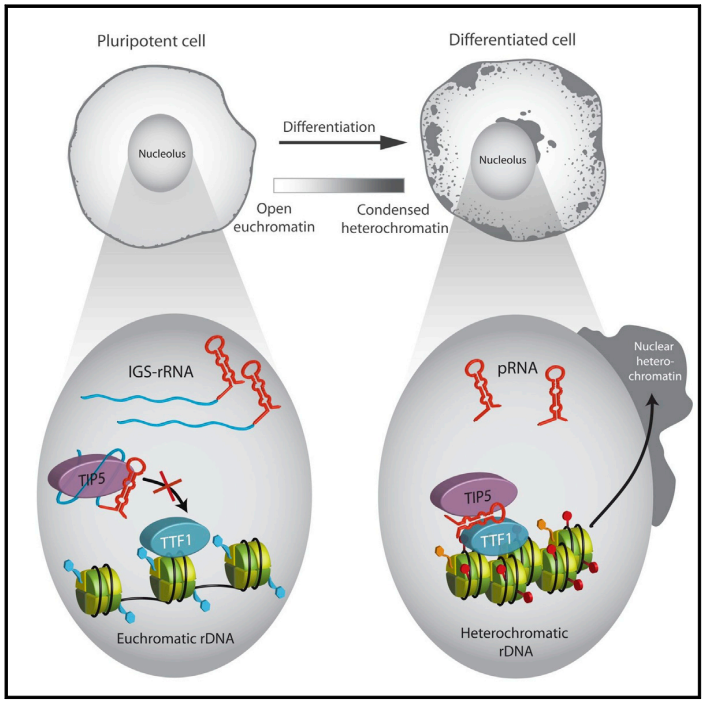
Authors:	**Sergio Leone** & Raffaella Santoro
Journal:	FEBS *Letters*
DOI:	10.1002/1873-3468.12308
External link:	http://onlinelibrary.wiley.com/doi/10.1002/1873-3468.12308/full
Contribution:	S.L. contributed to all figures. S.L. and R.S. wrote the manuscript.

Article

Cell Stem Cell

lncRNA Maturation to Initiate Heterochromatin Formation in the Nucleolus Is Required for Exit from Pluripotency in ESCs

Graphical Abstract



Highlights

rRNA genes (rDNA) acquire heterochromatin during ESC differentiation

Maturation of the lncRNA pRNA is required to establish rDNA heterochromatin

rDNA heterochromatin initiates heterochromatinization of ESC genomes

Inhibition of rDNA heterochromatin prevents ESC differentiation

Authors

Nataša Savić, Dominik Bär, ..., Paolo Cinelli, Raffaella Santoro

Correspondence

raffaella.santoro@vetbio.uzh.ch

In Brief

Savić et al. reveal that heterochromatin condensation in the nucleolus, where ribosomal genes are transcribed, triggers remodeling of the global open ESC chromatin into a highly condensed heterochromatic structure and that this mechanism is required for exit from pluripotency.

Savić et al., 2014, Cell Stem Cell 15, 720–734
December 4, 2014 ©2014 Elsevier Inc.
http://dx.doi.org/10.1016/j.stem.2014.10.005

CellPress

Cell Stem Cell
Article

lncRNA Maturation to Initiate Heterochromatin Formation in the Nucleolus Is Required for Exit from Pluripotency in ESCs

Nataša Savić,[1,2] Dominik Bär,[1] Sergio Leone,[1,2] Sandra C. Frommel,[1,2] Fabienne A. Weber,[2,3] Eva Vollenweider,[1,2] Elena Ferrari,[1] Urs Ziegler,[4] Andres Kaech,[4] Olga Shakhova,[5] Paolo Cinelli,[3,6,7] and Raffaella Santoro[1,6,*]
[1]Institute of Veterinary Biochemistry and Molecular Biology, University of Zurich, 8057 Zurich, Switzerland
[2]Molecular Life Science Program, Life Science Zurich Graduate School, University of Zurich, 8057 Zurich, Switzerland
[3]Institute of Laboratory Animal Science, University of Zurich, 8057 Zurich, Switzerland
[4]Center for Microscopy and Image Analysis, University of Zurich, 8057 Zurich, Switzerland
[5]Department of Oncology, University Hospital Zurich, 8952 Schlieren, Switzerland
[6]Center for Applied Biotechnology and Molecular Medicine, University of Zurich, 8057 Zurich, Switzerland
[7]Division of Trauma Surgery, Center for Clinical Research, University Hospital Zurich, 8091 Zurich, Switzerland
*Correspondence: raffaella.santoro@vetbio.uzh.ch
http://dx.doi.org/10.1016/j.stem.2014.10.005

SUMMARY

The open chromatin of embryonic stem cells (ESCs) condenses into repressive heterochromatin as cells exit the pluripotent state. How the 3D genome organization is orchestrated and implicated in pluripotency and lineage specification is not understood. Here, we find that maturation of the long noncoding RNA (lncRNA) pRNA is required for establishment of heterochromatin at ribosomal RNA genes, the genetic component of nucleoli, and this process is inactivated in pluripotent ESCs. By using mature pRNA to tether heterochromatin at nucleoli of ESCs, we find that localized heterochromatin condensation of ribosomal RNA genes initiates establishment of highly condensed chromatin structures outside of the nucleolus. Moreover, we reveal that formation of such highly condensed, transcriptionally repressed heterochromatin promotes transcriptional activation of differentiation genes and loss of pluripotency. Our findings unravel the nucleolus as an active regulator of chromatin plasticity and pluripotency and challenge current views on heterochromatin regulation and function in ESCs.

INTRODUCTION

The spatiotemporal organization of the genome has been recognized as an additional regulatory layer of chromatin, important for gene regulation and transcriptional competence (Gonzalez-Sandoval et al., 2013; Splinter and de Laat, 2011). Pluripotent stem cells such as embryonic stem cells (ESCs) are integral to the study of genome organization (Gorkin et al., 2014). Although ESCs organize their chromosomes into topological-associating domains that are largely invariant between cell types (Dixon et al., 2012; Nora et al., 2012), chromatin is generally less condensed and largely devoid of compact heterochromatin blocks compared to lineage-committed cells (Efroni et al., 2008; Jørgensen et al., 2007; Melcer et al., 2012; Meshorer et al., 2006). While ESC chromatin fibers occupy the entire nuclear volume, the highly compacted chromatin of differentiated cells is organized into discrete domains leading to large regions of the nucleus devoid of DNA (Fussner et al., 2010). Transcriptionally inactive chromatin in ESCs is unusually disorganized and tends to participate in fewer specific long-range interactions than in differentiated cells (de Wit et al., 2013). These results are consistent with a chromatin conformation that is particularly malleable and transcriptionally permissive in pluripotent cells and that may allow maintenance of a plastic state for the different transcriptional programs required for lineage specification (de Wit et al., 2013; Gaspar-Maia et al., 2011; Gorkin et al., 2014). Upon ESC differentiation, large-scale genome silencing takes place and ESC chromatin undergoes structural remodelling toward a highly condensed heterochromatic and transcriptionally repressed form (Bhattacharya et al., 2009; Meshorer and Misteli, 2006). These changes are also accompanied by alterations of nuclear architecture such as formation of large organized chromatin regions enriched in the heterochromatic and repressive histone modification H3K9 methylation (termed LOCKs) (Wen et al., 2009), maturation and compaction of constitutive heterochromatin (such as centric and pericentric repeats) and clustering of highly condensed heterochromatin either at the nucleolus or at the nuclear periphery (Aoto et al., 2006; Bártová et al., 2008a, 2008b; Efroni et al., 2008). However, how ESCs mediate the switch from a lower to a higher order chromatin structure remains elusive and calls for studies aimed at understanding the mechanism and function of this process.

An important component of nuclear architecture is the nucleolus, the compartment where transcription of hundreds of ribosomal RNA (rRNA) genes, rRNA processing, and ribosome subunit assembly take place (Moss and Stefanovsky, 2002). Clustering of highly condensed heterochromatin at nucleoli is a phenomenon known to occur in all somatic cells, yet neither the factors involved nor their physiological relevance is understood. Previous studies have however started to define a functional link between nuclear heterochromatin positioned in proximity to nucleoli and rRNA genes (rDNA), the genetic

component of the nucleolus. The nucleolar repressor factor TIP5 (TTF1-interacting protein 5, also known as BAZ2A) is required for heterochromatin formation of a fraction of rRNA genes through association with the long noncoding (lnc)RNA pRNA, DNA methyltransferases (DNMT1, DNMT3b), histone deacetylase HDAC1, and poly(ADP-ribose)-polymerase-1 (PARP1) (Guetg et al., 2012; Mayer et al., 2006; Santoro et al., 2002, 2010; Zhou et al., 2002). pRNA is a 250–300 nucleotide transcript corresponding to main rDNA promoter sequences and originates from processing of the 2 kb long IGS-rRNA (intergenic spacer rRNA) whose synthesis is driven by an alternative rDNA promoter located upstream of the main rDNA promoter (Mayer et al., 2006; Santoro et al., 2010). TIP5-pRNA interaction is necessary to form rDNA heterochromatin by mediating TIP5 nucleolar retention and association with rDNA and PARP1 (Guetg et al., 2012; Mayer et al., 2006). Depletion of TIP5 reduces silent epigenetic marks at rDNA and heterochromatic centric and pericentric repeats, and abrogates formation of condensed heterochromatic structures within and in proximity to the nucleolus (Guetg et al., 2010). Strikingly, this structural organization closely resembles the open chromatin of ESCs, prompting us to investigate if the chromatin state of the nucleolus regulates ESC chromatin plasticity and commitment to specific lineages.

RESULTS

Establishment of rDNA Heterochromatin Occurs during ESC Differentiation

We profiled the epigenetic state of the nucleolus, at rDNA, in ESCs and during differentiation into neural progenitor cells (NPCs) that are *Pax-6*, *Nestin*, and brain lipid-binding protein (*BLBP*) positive and do not express the pluripotency factor Nanog (Figure 1A; Figure S1A available online). Previous work showed that methylation of the two unique CpG dinucleotides at the mouse rDNA promoter distinguishes heterochromatic and silent rRNA genes from euchromatic, transcriptionally active rDNA (Santoro and Grummt, 2001). We quantified silent rDNA by measuring CpG methylation at rDNA promoter of ESCs, NPCs, and mouse somatic cells from brain tissue using HpaII digestion followed by qPCR, a method that accurately quantifies the amounts of silent rDNA (Santoro et al., 2002). Consistent with a previous bisulfite analysis (Schlesinger et al., 2009), rDNA promoter in ESCs displays very low meCpG levels, confirming the accuracy of our method. After 8 days of differentiation, a fraction of rRNA genes (25%–30%) acquired CpG methylation at levels comparable to brain tissue (Figure 1B). Similar results were obtained with a different ESC line and differentiation protocol (Figure S1B), indicating that rDNA is de novo methylated during early ESC differentiation. Consistent with these results, rDNA transcription levels were similar in NPCs and mouse fibroblast NIH 3T3 cells and lower than in ESCs (Figure 1C). Remarkably, rDNA methylation in induced pluripotent stem cells (iPSC) decreased to about one half when compared to the original fibroblasts (Figure 1D), implying a link between cell pluripotency and rDNA methylation levels. Upon differentiation, heterochromatin-related histone modifications H3K9me2, H3K9me3, and H3K27me3 increased at rDNA promoter and coding regions (Figure 1E; Figure S1C). In contrast, active histone modifications such as H3K4me2 and H3K4me3 were not greatly affected (Figure S1D). Consistent with previous reports, major and minor satellite repeats that compose centric and pericentric heterochromatin increased H3K9me3 levels during differentiation (Martens et al., 2005; Meshorer et al., 2006; Wong et al., 2009), whereas H3K9me2 occupancy was not greatly affected. These changes were accompanied by a reduction of major and minor satellite transcripts (Figure S1E), which are normally repressed in differentiated cells (Efroni et al., 2008). We conclude that formation of rDNA heterochromatin takes place during ESC-NPC transition and timely coincides with the switch to a higher condensed heterochromatic form of centric and pericentric repeats.

TIP5 Is Recruited to rDNA during ESC Differentiation

To determine how rDNA heterochromatin is established during ESC differentiation, we measured the association of upstream binding factor UBF, an essential rDNA transcription factor that exclusively binds to unmethylated, euchromatic rRNA genes, and of TIP5, which associates with methylated silent rDNA (Santoro and Grummt, 2001; Santoro et al., 2002). UBF occupancy at rDNA was lower in NPCs than in ESCs (Figure 1F), a further indication that the number of euchromatic active rRNA genes decreases during differentiation. In contrast, TIP5 binds to rDNA only in NPCs but not in ESCs (Figure 1F). Similarly, PARP1, previously shown to interact with TIP5 and implicated in the formation of rDNA heterochromatin (Guetg et al., 2012), increases its association with the rDNA promoter in NPCs. Thus, establishment of rDNA heterochromatin in ESC-NPC transcription is accompanied by a decrease in the association of factors specific to active genes and an increase in the binding of components of the rDNA silencing machinery.

TIP5 protein and mRNA levels were higher in ESCs than in NPCs (Figure 1G), implying that the lack of rDNA heterochromatin in ESC is independent of TIP5 amounts. We then analyzed the TIP5 cellular localization (Figure 1H; Figure S2). Consistent with previous results, TIP5 was exclusively localized within nucleoli of somatic MEFs, as indicated by the colocalization with the nucleolar protein UBF (Strohner et al., 2001). In contrast, the cellular localization of TIP5 in ESCs was predominantly nucleoplasmic and often excluded from the nucleoli. Upon differentiation, TIP5 was drastically reduced in the nucleoplasm and exclusively localized within the nucleoli of NPCs, showing a cellular localization that is characteristic of somatic cells (Figure 1H; Figure S2). We conclude that TIP5 association with rDNA is impaired in ESCs and its recruitment to rDNA is achieved upon ESC differentiation.

Processing of pRNA Mediates Formation of rDNA Heterochromatin

We reasoned that impairment of TIP5 binding to rDNA might be responsible for the lack of rDNA heterochromatin in ESCs. Previous studies implicated the lncRNA pRNA in TIP5 nucleolar retention and association with rDNA (Mayer et al., 2006). pRNA is a 250–300 nucleotide transcript corresponding to main rDNA promoter sequences and derives from processing of the 2 kb long IGS-rRNA (Mayer et al., 2006; Santoro et al., 2010) (Figure 2A). Measurements of pRNA sequences in ESCs and at different times of differentiation did not reveal remarkable differences between ESCs and NPCs (Figure 2B). However, this approach does not allow distinguishing between IGS-rRNA and mature pRNA.

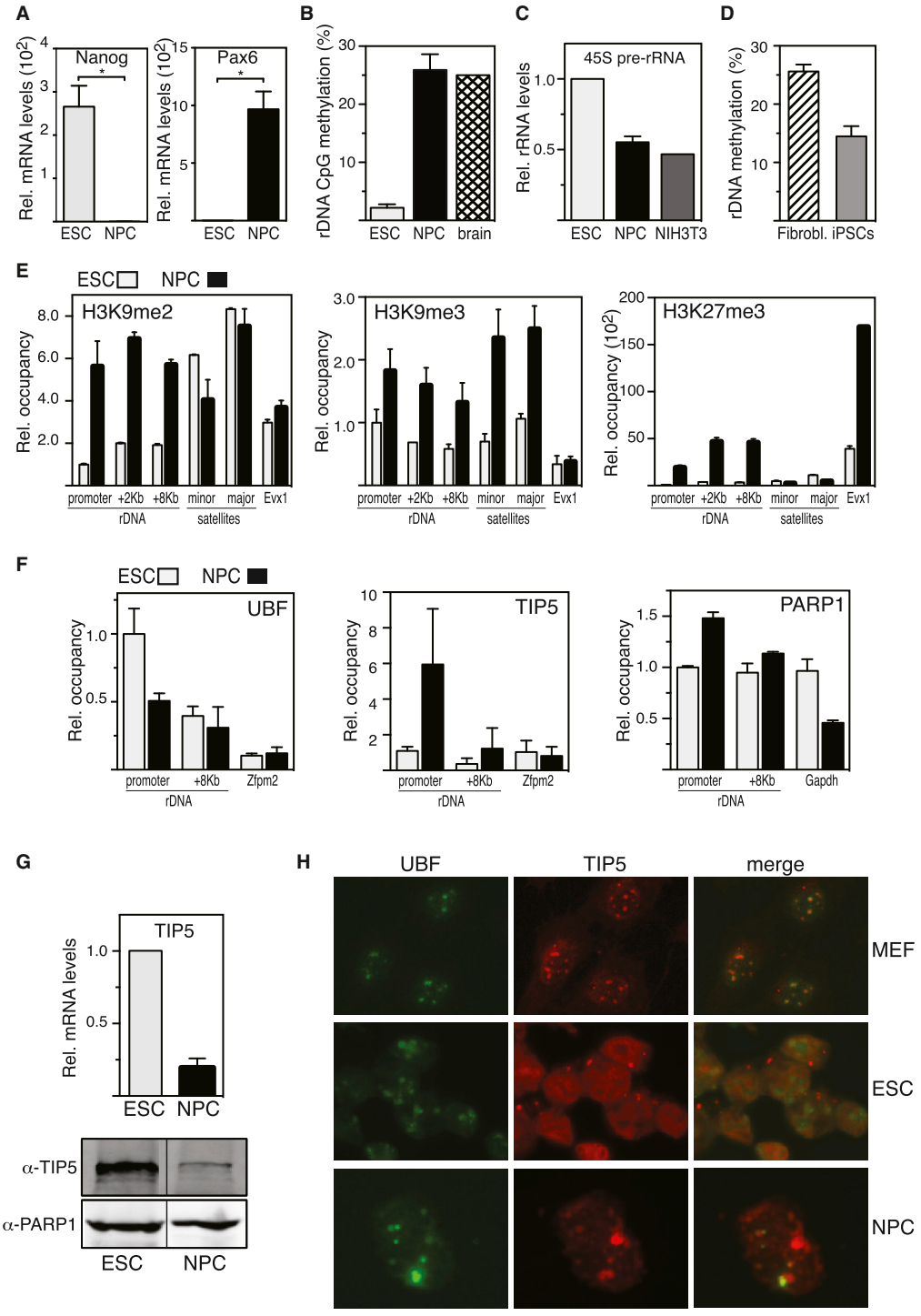


Figure 1. Establishment of rDNA Heterochromatin Occurs during ESC Differentiation and Correlates with the Recruitment of TIP5 to rDNA

(A) qRT-PCR. Nanog and Pax6 mRNA levels in ESCs and NPCs. Data were normalized to Rps12 mRNA.

(B) CpG methylation levels at rDNA promoter in ESCs, NPCs, and mouse brain tissues.

(legend continued on next page)

Quantification of 5′- and internal IGS-rRNA regions determined that unprocessed transcript levels were higher in ESCs than in NPCs (Figure 2B), suggesting that IGS-rRNA processing is less efficient in ESCs than in NPCs. To support these results, we measured ectopic IGS-rRNA and pRNA derived from an IGS-rRNA reporter gene plasmid that was transfected in NIH 3T3 cells, proficient for IGS-rRNA processing (Santoro et al., 2010), and in ESCs. Whereas ectopic IGS-rRNA was efficiently processed in NIH 3T3 cells (80%), maturation of pRNA was strongly reduced in ESCs (Figure 2C). Taken together, these results indicate that IGS-rRNA is not efficiently processed in ESCs and is thus less abundant in NPCs than in ESCs.

To test whether the lack of IGS-rRNA processing is the determinant that impairs formation of rDNA heterochromatin in ESCs, we transfected in vitro synthesized mature pRNA in ESCs and monitored TIP5 cellular localization, rDNA methylation, rDNA transcription, and H3K9me2 and H3K9me3 levels at rDNA (Figure 3). In ESCs transfected with pRNA, TIP5 decreased in the nucleoplasm and accumulated within nucleoli, as indicated by the colocalization with the nucleolar protein UBF (Figure 3A; Figure S3). Remarkably, the addition of pRNA in ESCs induced formation of heterochromatic rDNA as demonstrated by the reduction of rDNA transcription and the increase of both H3K9me2 and CpG methylation levels at the rDNA promoter (Figures 3B–3D; Figure S4). The modest increase in CpG methylation (from 1.7% to 4.4%) can also be attributed to the 2i conditions, recently described to lead to pronounced reduction in DNA methylation due to the downregulation of the de novo methyltransferases DNMT3a and DNMT3b (Leitch et al., 2013). Consistent with previous studies showing that TIP5 mediates dimethylation but not trimethylation of H3K9 at rRNA genes (Guetg et al., 2010; Santoro and Grummt, 2005), ectopic pRNA did not increase H3K9me3 levels at rDNA (Figure 3D; Figure S4). These results indicate that addition of pRNA in ESCs is sufficient to guide TIP5 to rDNA and to establish rDNA heterochromatin.

To determine how pRNA guides TIP5 to rDNA, we mutated pRNA sequences that were previously implicated in rDNA methylation and TIP5 association in somatic cells (Mayer et al., 2008; Schmitz et al., 2010). We mutated the T_0 element (pRNAΔT_0) that was previously described to form dsDNA:RNA triplex, a structure implicated in de novo rDNA methylation through recruitment of DNMT3b (Schmitz et al., 2010) and proposed as guiding module for TIP5 targeting to rDNA (Bierhoff et al., 2013). Similarly to wild-type pRNA, pRNAΔT_0 induced recruitment of TIP5 to nucleoli, promoted rDNA methylation and reduced rDNA transcription (Figures 3A–3C; Figure S3). Thus, pRNA-mediated nucleolar targeting of TIP5 and establishment of rDNA heterochromatin formation in ESCs is not mediated by rDNA:pRNA triplex. Accordingly, replacement of the 5′-pRNA region, including T_0 element, (hybrid, Control-pRNA) induced TIP5 nucleolar localization (Figure 3E; Figure S3). In contrast, replacement of 3′-pRNA sequences (hybrid, pRNA-Control), important for stem loop structure formation and the association with TIP5 in vitro (Mayer et al., 2008), impaired nucleolar localization of TIP5. Remarkably, point mutations that disrupt the stem loop structure (pRNA loop destroyed) were not efficient in recruiting TIP5 to the nucleoli whereas a compensatory mutation allowing hairpin formation did (pRNA loop recovered). Together, these results indicate that pRNA guides TIP5 to rDNA in *trans* through the hairpin structure and that addition of mature pRNA to ESCs is sufficient to establish rDNA heterochromatin. We conclude that the impairment of IGS-rRNA processing that abrogates formation of mature pRNA is the major determinant causing the euchromatic state of all rRNA genes in ESCs.

TIP5-TTF1 Association Is Mediated by pRNA and Impaired by IGS-rRNA

To determine why IGS-rRNA is unable to promote recruitment of TIP5 to rDNA, we determined whether TIP5 binds to IGS-rRNA using EMSA competition assays. Consistent with previous results, pRNA had a higher affinity for TIP5 compared to control RNA (Mayer et al., 2006) (Figure 4A). Surprisingly, TIP5 associates better with IGS-rRNA than with pRNA. To determine whether TIP5 preferentially associates with other IGS-rRNA sequences than pRNA, we analyzed TIP5 binding to IGS-rRNA sequences located upstream of the pRNA region. Spacer and enhancer RNA associate with TIP5 much less efficiently than pRNA (Figure S5A), suggesting that TIP5 binds to IGS-rRNA through the pRNA sequence and that upstream sequences might stabilize the complex through weak interactions. Together, these results indicate that TIP5 binds to IGS-rRNA and suggest that impairment of TIP5 recruitment to rDNA in ESCs might depend on the context of this interaction.

The requirement of the pRNA loop structure for nucleolar targeting of TIP5 (Figure 3) let us hypothesize that pRNA binding to TIP5 might favor the association with a docking protein for the recruitment to rDNA promoter and that IGS-rRNA might hinder this process. One important TIP5 interacting protein is the transcription terminator factor TTF1 (Németh et al., 2004; Strohner et al., 2001). TTF1 is a nucleolar protein that binds to terminator (T) elements, including the T_0 sequences at rDNA promoter, and is implicated in several rDNA regulatory processes (Evers and Grummt, 1995; Gerber et al., 1997; Längst et al., 1997). The association of TIP5 with TTF1 and its dependency on TTF1 for rDNA promoter binding proposed that TTF1 recruits TIP5 to rDNA (Németh et al., 2004; Santoro and Grummt, 2005; Strohner et al., 2001). However, whether and how pRNA is implicated in this process has so far not been investigated. TTF1 binds to RNA (Figure 4B), forming high-molecular-weight complexes. However, in contrast to TIP5, TTF1 did not display any

(C) qRT-PCR. rRNA transcription (45S pre-rRNA levels) in ESCs, NPCs, and NIH 3T3. Data were normalized to Rps12 mRNA.
(D) CpG methylation levels at rDNA promoter in mouse fibroblasts and iPSCs.
(E and F) ChIP. H3K9me2, H3K9me3 and H3K27me3, UBF, TIP5, and PARP1 occupancy. *Evx1*, *Zfpm2*, and *Gapdh* represent control genes. Data of two independent experiments were normalized to input and rDNA promoter value in ESCs.
(G) Tip5 mRNA (qRT-PCR) and protein levels (immunoblot) of ESCs and NPCs. Data were normalized to *Rps12* mRNA or PARP1 protein levels. Protein level titration was loaded and only the lanes with same protein amounts are shown.
(H) TIP5 cellular localization in MEFs, ESCs, and NPCs after 8 days of differentiation by immunofluorescence. Nucleoli are depicted by UBF signal.
All error bars represent the SD of two (when indicated) or three independent experiments. See also Figures S1 and S2.

Cell Stem Cell
The Nucleolus Regulates ESC Chromatin

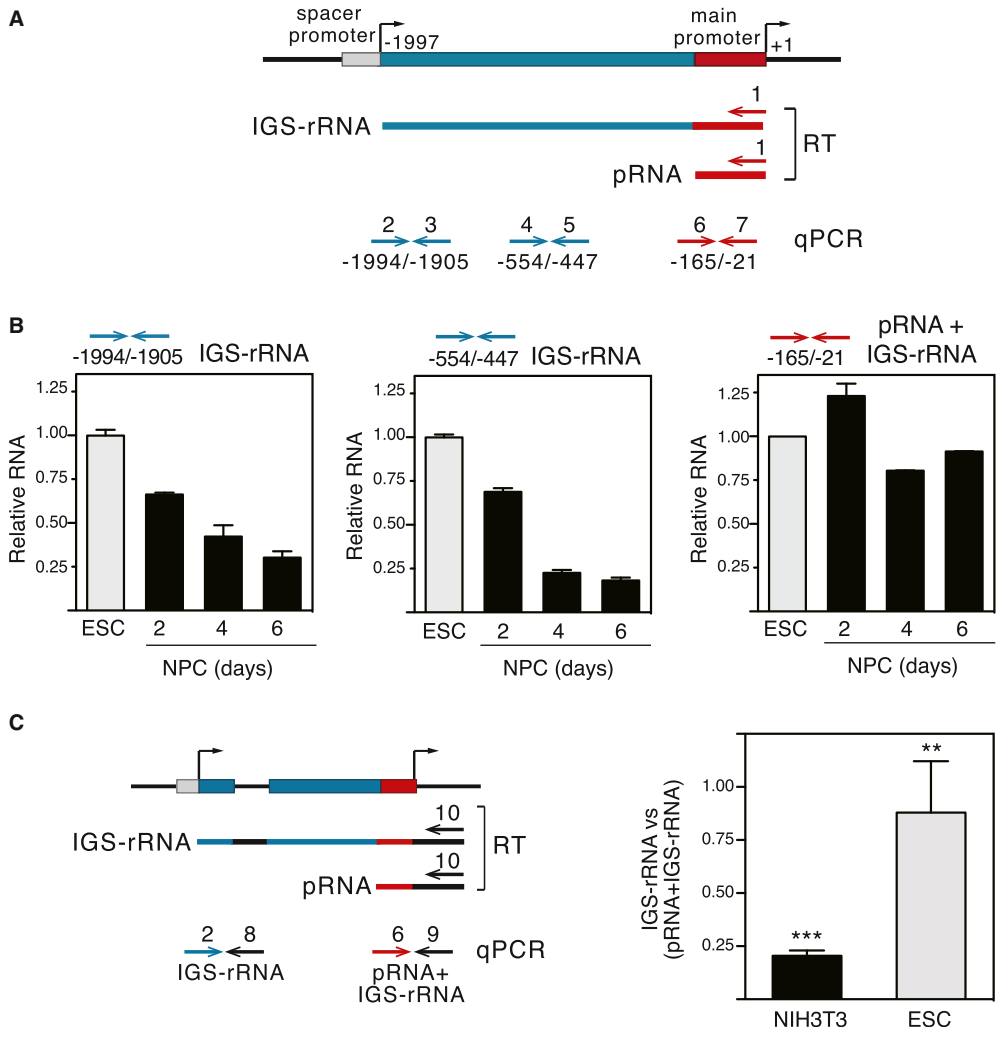


Figure 2. IGS-rRNA Is Not Efficiently Processed in ESCs

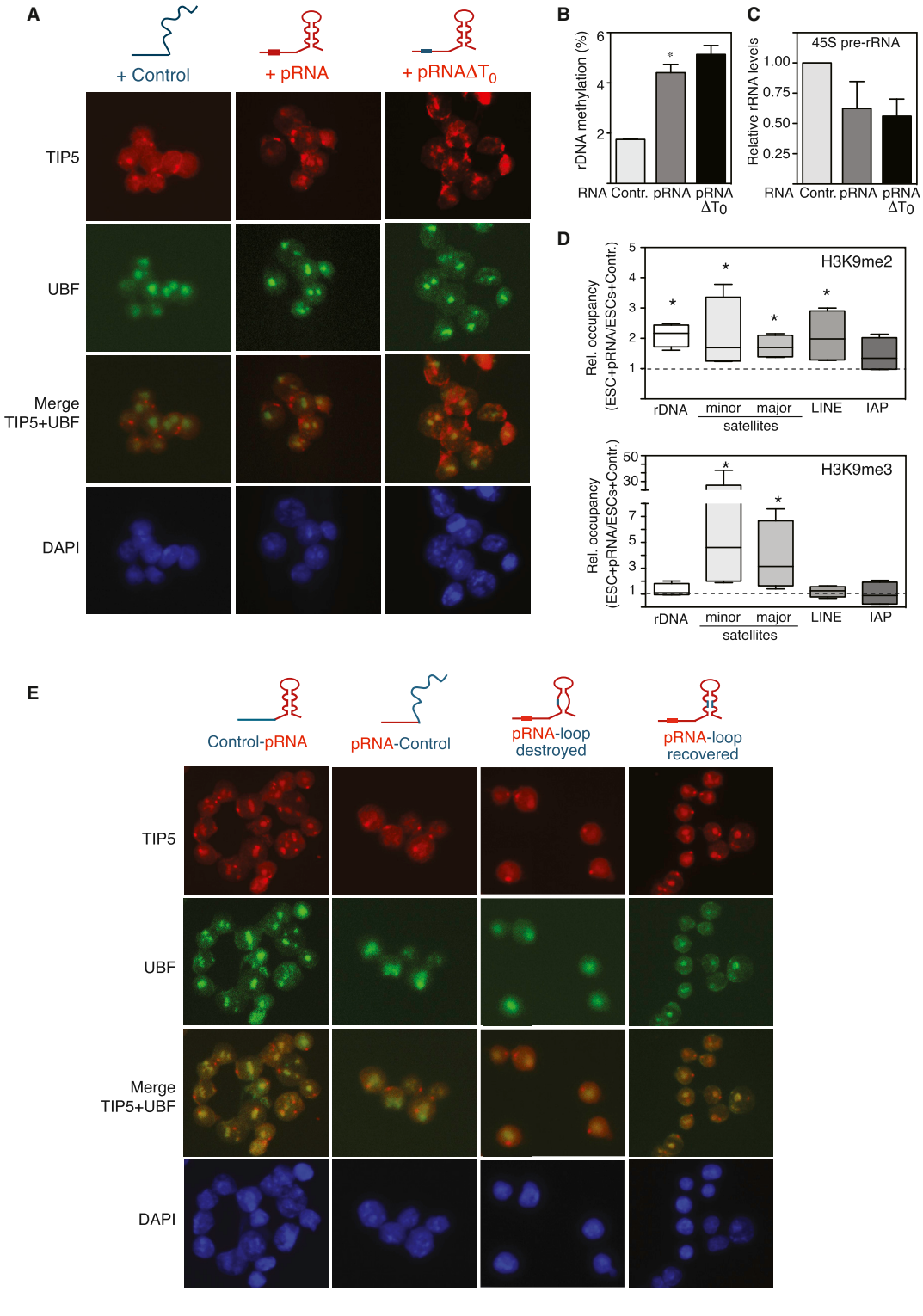
(A) Schema representing the mouse 5′-rDNA organization: Spacer promoter (gray), intergenic spacer region (blue), rDNA main promoter (red), and transcription start sites of IGS-rRNA (−1997) and 45S pre-rRNA (+1). Arrows represent primers used to perform RT (1; −20/−1) and to quantitatively amplify the indicated rDNA sequences (2–7).

(B) qRT-PCR. Levels IGS-rRNA and pRNA sequences of ESCs and NPCs (from days 2 to 6, from the beginning of differentiation). Data of two independent experiments were normalized to *Rps12* mRNA and to ESC values.

(C) Schema depicts the IGS-rRNA reporter plasmid. Black arrows represent primers used to perform RT (10) and to amplify plasmid sequences (8 and 9). Blue and red arrows (2 and 6) indicate primers hybridizing to rRNA sequences as described in (A). NIH 3T3 and ES cells were transfected with IGS-rRNA reporter plasmid. Data from three experiments are represented as values of amplifications with primers 2 and 8 (IGS-rRNA) normalized to amplifications with primers 6 and 9 (IGS-rRNA$^+$pRNA).

preferential binding to pRNA sequences (Figure 4C). In ESCs, TTF1 is bound to rDNA promoter and is localized within nucleoli as in differentiated cells (Figures S5B and S5C). To determine whether the association of TIP5 with TTF1 is regulated by pRNA or IGS-rRNA, we performed pull-down assays using purified RNA-free His-tagged TTF1 (aa.1-210, containing TIP5-interacting region) and GST-tagged-TIP5 (aa. 332-723, comprising the RNA- and TTF1-interacting regions) (Mayer et al., 2006; Németh et al., 2004). Immobilized GST-TIP5$_{332–723}$ was incubated with no RNA, or with equivalent moles of RNAs and analyzed for the interaction with His-TTF1$_{1–210}$. In the absence of RNA and in the presence of RNA control, TIP5 and TTF1 did not associate (Figure 4D). In contrast, TIP5-pRNA complexes displayed a strong interaction with TTF1, indicating that pRNA is required for TIP5-TTF1 association. Remarkably, TIP5 bound to IGS-rRNA did not interact with TTF1. Consistent with the role of 3′-pRNA sequences in TIP5 nucleolar targeting, this region was sufficient for TIP5-TTF1 interaction while the 5′-pRNA

Cell Stem Cell
The Nucleolus Regulates ESC Chromatin



(legend on next page)

region was not. We conclude that pRNA mediates the association of TIP5 with TTF1 and that unprocessed IGS-rRNA prevents this interaction. Based on these results, we propose that the unprocessed IGS-rRNA in ESCs abolishes the interaction of TIP5 with TTF1, thus preventing TIP5 targeting to rDNA and nucleolar heterochromatin formation.

The Epigenetic State of Nucleolar Chromatin Affects ESC Chromatin and Pluripotency

Our previous studies showed that somatic cells depleted of TIP5 reduced silent epigenetic modifications at rDNA and at major and minor satellites and lack condensed heterochromatin adjacent to nucleoli (Guetg et al., 2010). Further structural alterations observed upon TIP5 knockdown, such as enlargement of nucleolar surfaces, reduction of nucleoli number, and formation of few decondensed DAPI-stained heterochromatic foci, resembled ESC chromatin organization (Figure 5A). To determine whether formation of rDNA heterochromatin affects the open chromatin of ESCs, we performed transmission electron microscopy to analyze and quantify the volume of heterochromatin associated to nucleoli in ESCs upon addition of pRNA (ESCs+pRNA) (Figure 5B; Figure S5D). Consistent with the loss of perinucleolar heterochromatin upon TIP5 knockdown in somatic cells (Guetg et al., 2010), ESCs+pRNA increased the amounts of condensed heterochromatin around nucleoli, displaying a structural organization like it is found in NPCs (Figure S5E). In line with these results, ESCs+pRNA increased H3K9me2 not only at rDNA, but also at major and minor repeats and at LINE elements while IAP transposons were not significantly affected (Figure 3D; Figure S4). H3K9me3 levels drastically increased at minor and major satellites while remaining unchanged at LINE L1 elements and IAP transposons and, as expected, at rDNA. Among the four performed experiments, we observed an inverse correlation between H3K9me2 and H3K9me3 levels at minor and major repeats (Figure S2), suggesting a two-step process that is initiated with H3K9me2 (the activity brought to rDNA by TIP5 (Santoro and Grummt, 2005)) and is further completed at major and minor satellite sequences with the establishment of H3K9me3. ESCs+pRNA increased the total amount of H3K9me2 to levels similar to those observed during ESC-NPC transition (Figure 5C). This result is in agreement with a previous work showing acquisition of large regions of H3K9 methylation during differentiation, which affects at least 30% of the genome (Wen et al., 2009). In contrast, the global H3K9me3 content was not altered in NPCs and ESCs+pRNA. Because terminally differentiated cells were previously described to contain elevated H3K9me3 levels (Hawkins et al., 2010), our results with NPCs most likely represent the epigenetic state at early differentiation time points. The elevated heterochromatic content of ESCs+pRNA was also accompanied by a reduction of minor satellites transcription as found during ESC differentiation (Figure S1E), whereas transcription of LINE and IAP elements was not greatly affected (Figure 5C). Taken together, these results suggest that tethering heterochromatin at rDNA via pRNA in ESCs initiates structural remodeling toward a highly condensed nuclear heterochromatin, a structure that ESCs normally acquire during differentiation.

We next addressed how the increased heterochromatic content mediated by pRNA affects ESC properties. ESCs+pRNA did not show alterations in important molecular features of the undifferentiated state such as cell proliferation, expression of the pluripotency genes *Oct4*, *Nanog*, and stage-specific embryonic antigen 1 (SSEA1), cell morphology, and alkaline phosphatase (AP) staining (Figures 6A–6D, and data not shown). To determine whether addition of pRNA affects pluripotency in vivo, we performed teratoma formation assays. ESCs+pRNA markedly decreased their capability to form teratoma compared to ESCs transfected with RNA control or with a pRNA mutant that is unable to recruit TIP5 to nucleoli (Figure 6E). Remarkably, teratomas derived from ESCs+pRNA showed a drastic reduced volume but displayed differentiation into all three germ layers (Figures S6A and S6B). Because ESCs were transiently transfected, we assume that teratomas obtained from ESCs+pRNA originated from untransfected cells, which is supported by the reduction in volume of these teratomas. To get insights into the loss of pluripotency, we analyzed transcription profiles of ESCs+pRNA and ESCs+RNA-control by RNaseq and found upregulation of 532 transcripts and downregulation of 509 transcripts in ESCs+pRNA (Table S1 available online). We carried out functional annotation analysis of transcripts whose levels were altered in ESCs+pRNA using DAVID tools (Huang et al., 2009) (Figure 6E; Table S1). The top eight gene ontology terms were all related to cell developmental and differentiation processes. Enrichment for these processes was particularly evident for transcripts upregulated in ESCs+pRNA, suggesting that addition of pRNA promotes expression of genes involved in cell differentiation and developmental processes. To exclude the possibility of pRNA off-target effects, we transfected a mutant pRNA unable to associate with TIP5 and analyzed transcription of *Btg3* and *Rdh10*, two genes upregulated in ESCs+pRNA and known to be implicated in neurogenesis and embryonic differentiation (Cammas et al., 2007; Yoshida et al., 1998). In contrast to pRNA, the pRNA mutant was unable to affect Btg3 and Rdh10 transcript levels (Figure S6C), indicating that upregulation of genes implicated in cell differentiation and the developmental processes is specifically mediated by a fully functional pRNA that is able to associate with TIP5, guide it to

Figure 3. Mature pRNA Is Required for the Establishment of rDNA Heterochromatin

(A) Mature pRNA induced recruitment of TIP5 to ESC nucleoli. Immunofluorescence with anti-TIP5 and anti-UBF antibodies in ESCs transfected with in vitro synthesized RNA control, pRNA, and pRNAΔT_0.

(B) Mature pRNA promotes rDNA promoter CpG methylation in ESCs. Error bars represent the SD of three independent experiments.

(C) Mature pRNA decreased rRNA synthesis in ESCs. rDNA transcription were measured by qRT-PCR and normalized to *Rps12* mRNA. Error bars indicate the SD of two independent experiments.

(D) Mature pRNA increased silent histone modifications at rDNA and centric-pericentric sequences of ESCs. Box-and-whisker plot of four independent ChIP experiments. Data are represented as bound over input in ESCs+pRNA normalized to values measured in ESCs+RNA control.

(E) pRNA loop structure mediates TIP5 recruitment to ESC nucleoli. Immunofluorescence with anti-TIP5 and anti-UBF antibodies of ESCs transfected with the indicated pRNA mutants.

See also Figure S3.

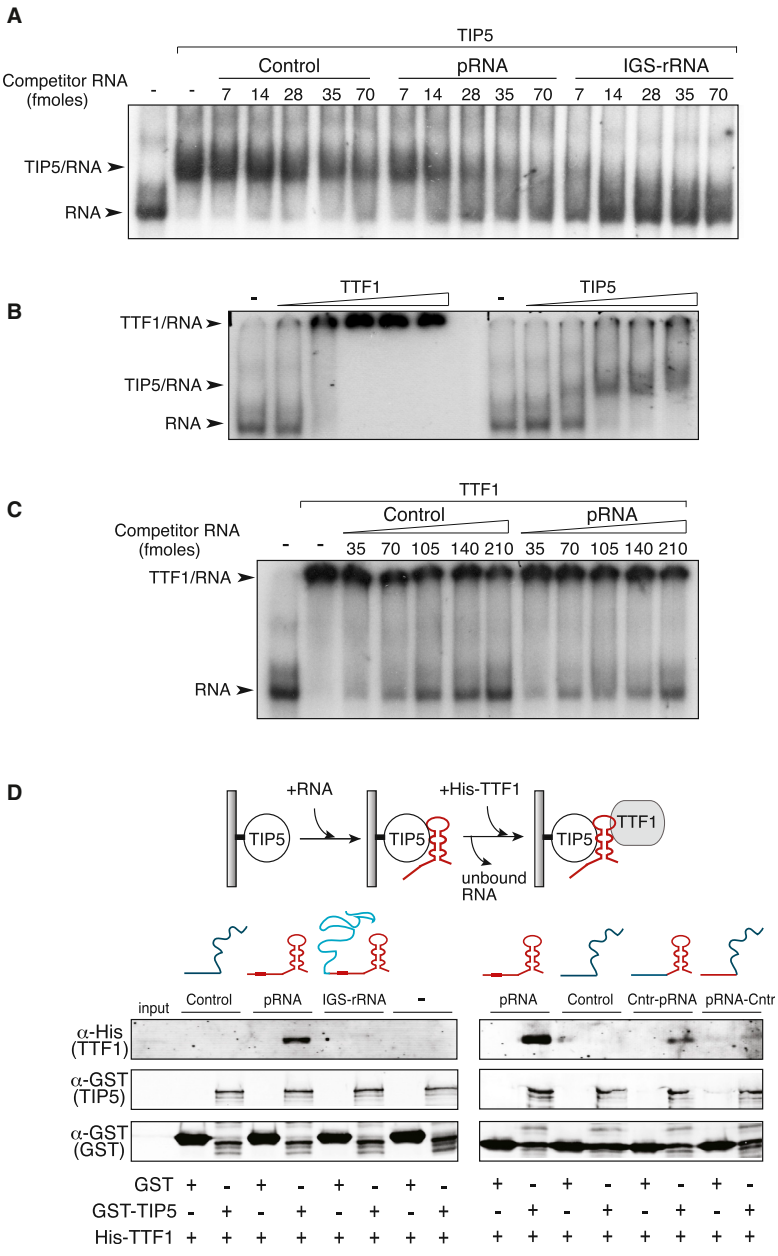


Figure 4. pRNA Mediates TIP5-TTF1 Interaction

(A) TIP5 binds to IGS-rRNA. Increasing equal moles of in vitro transcripts corresponding to control, pRNA, and IGS-rRNA sequences were used to compete for binding of $TIP5_{332–723}$ to radiolabelled run-off transcripts from pBluescript (MCS-RNA). RNA/protein complexes were analyzed by EMSA.

(B) TTF1 binds to RNA. Increasing equal moles of full-length TTF1 and $TIP5_{332–723}$ were analyzed for binding to radiolabeled MCS-RNA by EMSA.

(C) TTF1 does not show preferential binding to pRNA. Binding of full-length TTF1 to radiolabeled MCS-RNA was completed with increasing equal moles of RNA control and pRNA.

(D) TIP5-TTF1 interaction is mediated by pRNA and impaired by IGS-rRNA. Schema representing the GST-pull-down strategy used to analyze TIP5-TTF1 interaction in the presence of equivalent moles of RNAs. Pull-down assay. Bound proteins, GST, GST-TIP5, and His-TTF1, were detected with anti-GST and anti-His antibodies, respectively.

See also Figures S5A–S5C.

rDNA, and establish rDNA heterochromatin. Together, these results indicate that the elevated heterochromatic content induced by formation of nucleolar heterochromatin through pRNA impairs pluripotency and highlight the role of the nucleolus in the control of ESC chromatin plasticity that is required for the maintenance of the undifferentiated state.

To further explore the role of nucleolar chromatin in ESCs, we analyzed the differentiation capacity of cells depleted of TIP5. ESCs depleted of TIP5 by siRNA reduced TIP5 levels to 50% (Figures 7A and 7B). Upon TIP5 knockdown, ESCs proliferated less but displayed unaltered expression of *Oct4*, *Nanog*, and *Rex1*, exhibited the typical cell morphology of ESCs and were positive for AP staining (Figures 7B–7E). Similar results were obtained with other siRNA-TIP5 sequences (data not shown). To determine whether TIP5 depletion affects ESC differentiation, we treated an equal number of siRNA-control and -Tip5 treated ESCs with their respective siRNAs and induced monolayer differentiation upon withdrawal of 2i and LIF (Figure 7A). After 3 days, control cells displayed morphological structures typical of differentiated cells, whereas cells depleted of TIP5 underwent cell death and detached from the plate (Figures 7F–7H). The majority of the few siRNA-TIP5 cells that remained attached to plates showed ESC-like morphology and were positive for AP staining (Figures 7G and 7H; Figure S7). The effects observed with TIP5 depletion were specific for differentiated cells because ESCs double-treated with siRNA-Tip5 and cultured in 2i and LIF did not show any defect in viability (data not shown). These results indicate that TIP5 is required for ESC differentiation and suggest that the establishment of nucleolar heterochromatin is an event required for early differentiation.

DISCUSSION

The remodeling of the open and euchromatic genome structure of ESCs toward a highly condensed heterochromatic form

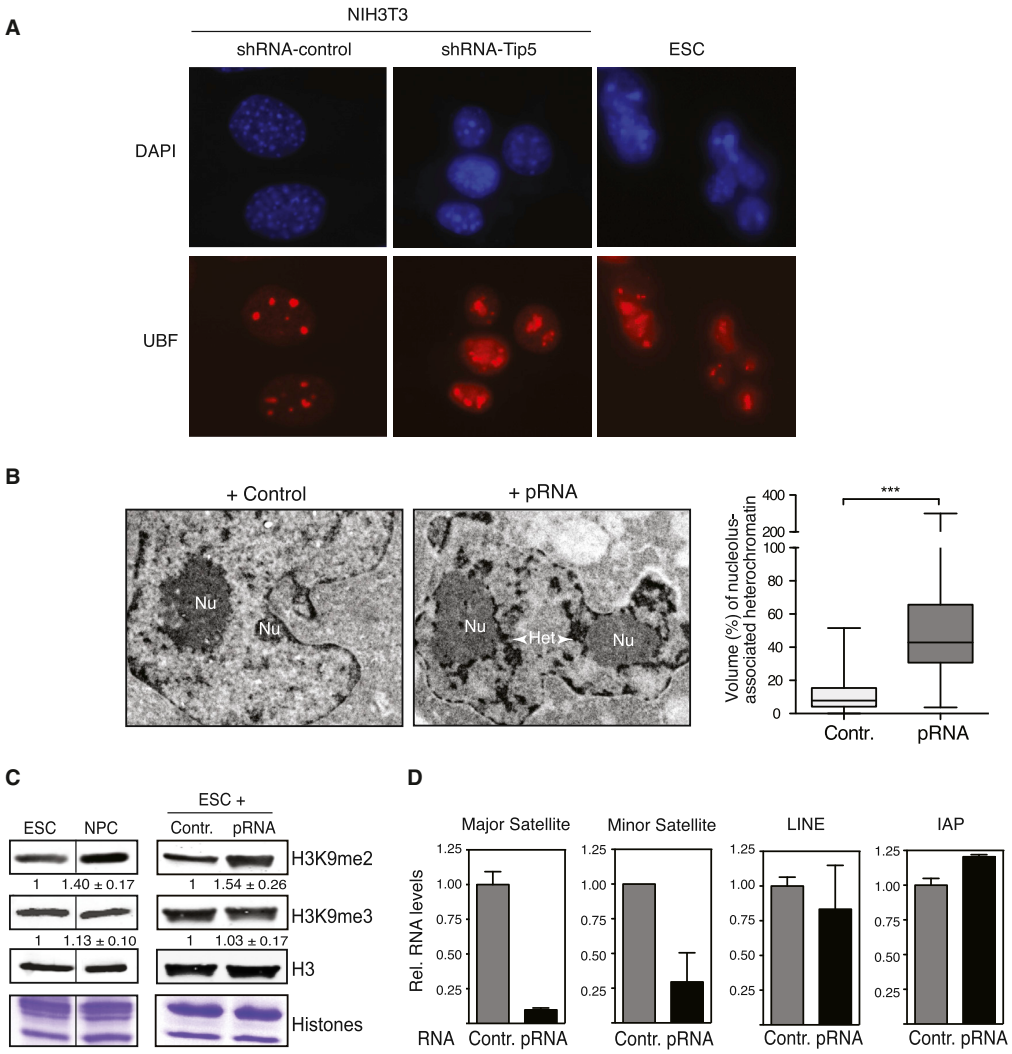


Figure 5. Mature pRNA Induces Global Remodelling toward Heterochromatic Structures

(A) Immunofluorescence showing heterochromatin (DAPI) and nucleoli (UBF) of NIH 3T3 cells stably expressing shRNA-control and -Tip5 sequences, and ESCs.

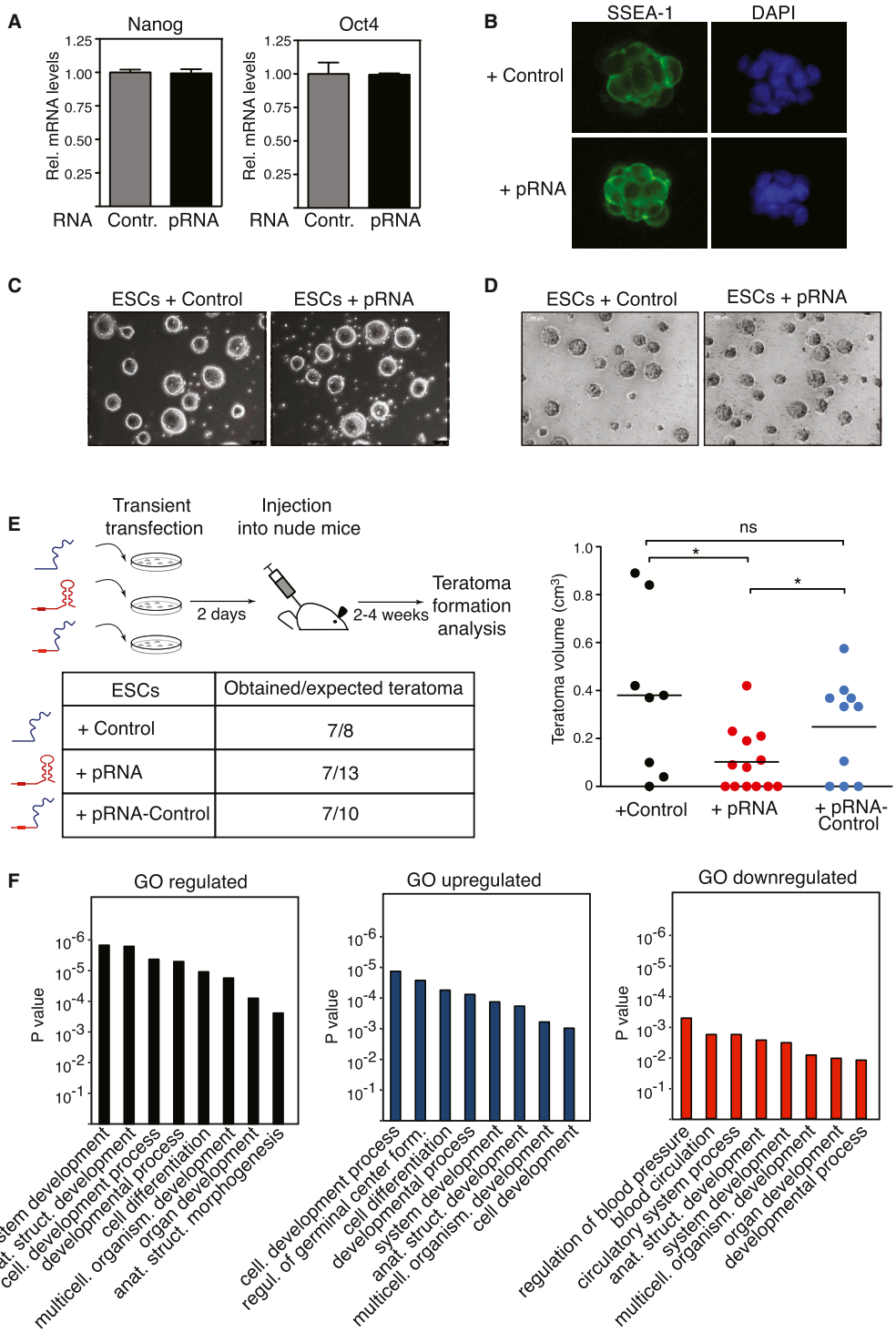
(B) Mature pRNA induced formation of condensed heterochromatin in ESCs. Transmission electron microscopy analysis. Representative images showing nucleolus-associated heterochromatin in ESCs+Control and ESCs+pRNA. The contrast procedure reveals in dark condensed heterochromatic structures (Het); 25–34 nucleoli of ESCs+RNA control and ESCs+pRNA of two independent experiments were selected at random. The volume of nucleolus-associated heterochromatin was expressed as a percentage of the volume of the nucleolus (Nu). Heteroscedastic two-tailed t test was used to compare the groups. See also Figures S5D and S5E.

(C) Immunoblot of H3K9me2, H3K9me3, and histone H3 of chromatin fractions of ESCs, NPCs, ESCs+RNA-control, and ESCs+pRNA. Values from three independent experiments were normalized to histone H3 levels. Protein level titration was loaded and only the lanes with same histone amounts are shown.

(D) qRT-PCR of major and minor satellite, LINE, and IAP retrotrasposon transcripts in ESCs+RNA-control and ESCs+pRNA. Values from two independent experiments were normalized to Rps12 mRNA.

characterizes the exit from pluripotency and the progression into differentiated states. ESC open chromatin correlates with a globally permissive transcriptional state contributing to the developmental plasticity and pluripotency of the ESC genome that has to have the ability to enter any distinct differentiation pathway (Gaspar-Maia et al., 2011). We determined here that the nucleolus is not only the place where ribosomes are produced, but it also plays a role in regulating genome architecture and pluripotency. Using mature pRNA as a mean to specifically tether heterochromatin at nucleoli of ESCs, we showed that the formation of heterochromatin at rRNA genes, the genetic component of nucleoli, has the ability to initiate the establishment of repressive chromatin structures at regions of the genome located outside of the nucleolus (Figure 7I). This process includes the formation of

Cell Stem Cell
The Nucleolus Regulates ESC Chromatin



(legend on next page)

highly condensed heterochromatic structures clustering in proximity to nucleoli as found in differentiated cells. Such architectural remodeling is also evidenced by a global increase in H3K9me2, maturation of heterochromatin at repetitive sequences, and their transcriptional repression. These are characteristic features of the switch between pluripotency and differentiation. Although we cannot define which is the first event that globally initiates the formation of heterochromatin at the exit from pluripotency and entry into differentiation, our results place the nucleolus as an important regulator of this process. The establishment of rDNA heterochromatin during differentiation timely coincides with the formation of highly condensed heterochromatic structures and LOCKs (Meshorer and Misteli, 2006; Wen et al., 2009). This was particularly evident for the maturation of constitutive heterochromatin at major and minor satellite repeats, which displayed the same timing as rDNA for the acquisition of histone repressive marks and transcriptional repression upon ESC differentiation.

While showing some of the molecular outlines of the undifferentiated cells, pRNA-mediated heterochromatic ESCs transcribe genes implicated in differentiation processes and are no longer pluripotent. This observation highlights the role of the euchromatic organization in ESC identity and suggests that nucleolar chromatin is an important regulator of the pluripotent state. Likewise, impairing the formation of rDNA heterochromatin by TIP5 knockdown inhibits ESC differentiation, suggesting that the establishment of nucleolar heterochromatin is a necessary step for the switch from a lower to a higher order chromatin structure and lineage commitment.

How does the nucleolus influence heterochromatin formation? Due to the proximity of rDNA and centromeric sequences at rDNA-bearing chromosomes (Dev et al., 1977; Kurihara et al., 1994), the effect of rDNA heterochromatin on chromatin structures at major and minor repeats might in part be explained through spreading mechanisms. However, centromeres of chromosomes not containing rRNA genes also associate with the nucleolus at a frequency more than expected for a random distribution (Carvalho et al., 2001), indicating the existence of alternative mechanisms that establish heterochromatin at sequences that are further away from rDNA loci. In this case, the establishment of rDNA heterochromatin might allow the formation of a nucleolar/perinucleolar compartment enriched in chromatin repressor complexes that would be attractive for genomic regions that need to be repressed. Anchoring of heterochromatin at the nucleolus might have similar function like the ones described for the nuclear periphery that is responsible for the integrity of mammalian heterochromatin (Pinheiro et al., 2012; Towbin et al., 2012). Consistent with this, genomic regions localized at the lamina (LADs) were shown to relocate after cell division either at the lamina or at the nucleolus (Kind et al., 2013), suggesting interchangeable roles in establishing and maintaining heterochromatic states.

This study also provides evidence that rRNA genes do not only function in synthesizing rRNA. Silent rRNA repeats, present in all somatic cells, maintain their heterochromatic state independently of transcriptional activity, and are stably propagated throughout the cell cycle (Conconi et al., 1989). Our results indicated that the epigenetic state of rRNA genes contributes to nuclear architecture and cellular functions such as pluripotency and differentiation by controlling the balance between heterochromatin and euchromatin. Interestingly, rDNA deletions in *Drosophila* result in reduced heterochromatin elsewhere in the genome and the extent of the rDNA deletion correlates with the loss of silencing in much the same manner as mutations in known protein heterochromatin components (Paredes and Maggert, 2009).

Very little is understood about how specific lncRNAs selectively seek out interaction sites in the genome, the nature of lncRNA-chromatin interactions, and their possible functional roles (Rinn and Chang, 2012). This work underlined the role of lncRNA in targeting epigenetic regulatory processes at specific genomic loci leading to the establishment of chromatin conformation patterns that ultimately result in the fine control of genes. We show that the regulation of pRNA precursor IGS-rRNA processing is a key determinant for the control of the epigenetic state at rDNA and propose that the processing represents a mean of lncRNA regulation to modulate distinct functions of the same lncRNA. We determined that IGS-rRNA processing is a developmentally regulated process and that its impairment in ESCs prevents recruitment of TIP5 to rDNA and formation of rDNA heterochromatin. Although the mechanisms that impair IGS-rRNA processing in ESCs are yet to be determined, our results demonstrated that mature pRNA is necessary to establish rDNA heterochromatin. We showed that pRNA-mediated targeting of TIP5 occurs through DNA-protein recognition rules. Whereas IGS-rRNA abolishes the association of TIP5 with TTF1, pRNA promotes this interaction that serves to guide the complex to the rDNA promoter and to establish nucleolar heterochromatin. Thus, the same lncRNA can prevent or promote protein complex assembly and its processing controls the switch between these functions. Based on these results, it would not be surprising if processing emerges as a more general mechanism of lncRNA regulation.

In summary, our data underline the contribution of chromatin structure in ESC pluripotency and differentiation potential and indicate that the nucleolus is a key regulator of nuclear architecture and chromatin structure, which serves to control cell pluripotency and lineage commitment.

Figure 6. pRNA Impairs ESC Pluripotency

(A) Nanog and Oct4 mRNA levels in ESCs+RNA control and ESCs+pRNA (qRT-PCR). Values from two independent experiments were normalized to *Rps12* mRNA.

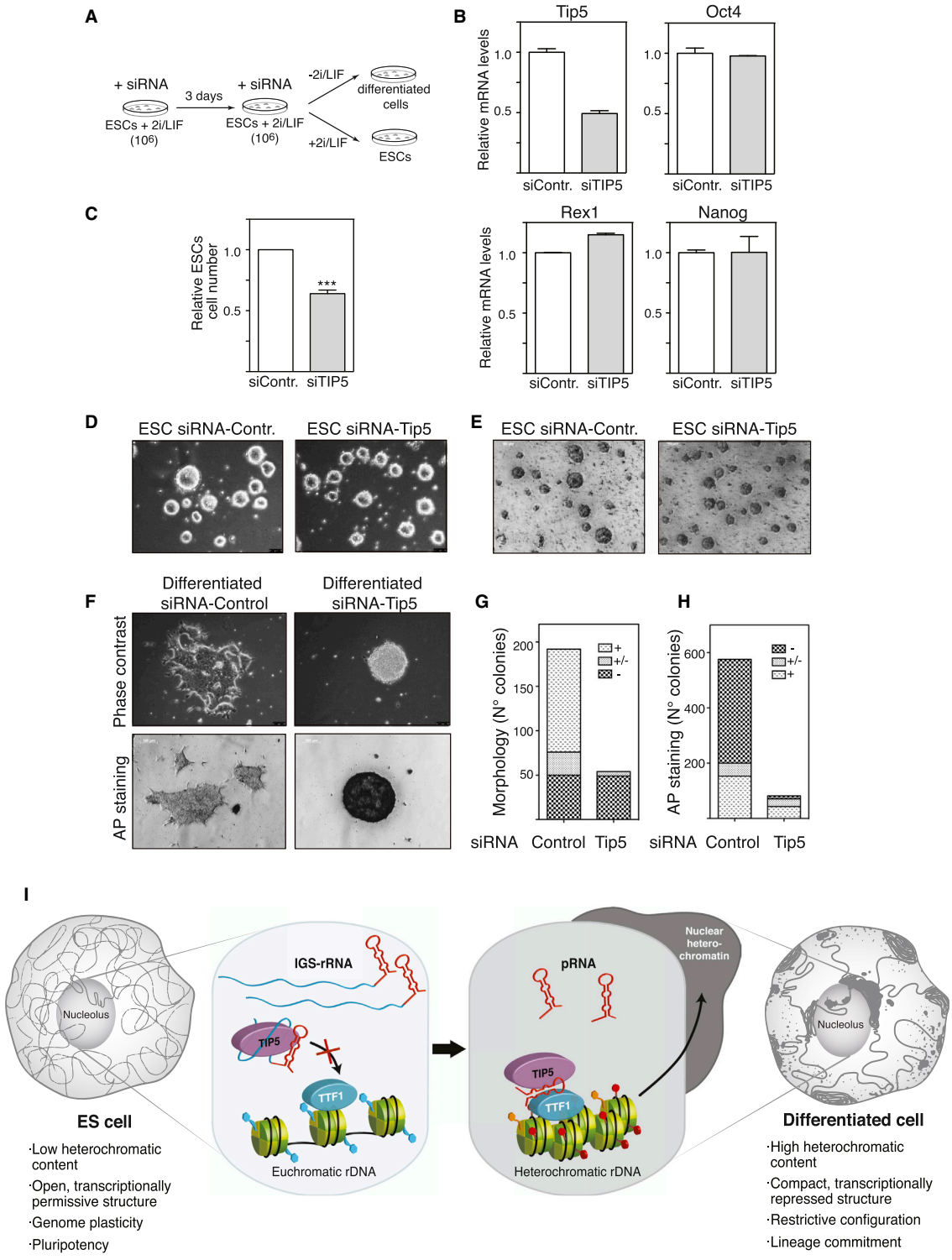
(B–D) (B) SSEA-1 immunostaining, (C) cell morphology, and (D) AP staining of ESCs+RNA control and ESCs+pRNA.

(E) ESCs+pRNA are not pluripotent. Efficiency of teratoma formation was assessed by the number of teratomas generated versus expected (injections) and by tumor volumes measured at the time when control animals were killed.

(F) Mature pRNA induced expression of genes implicated in cell differentiation and development. Top eight biological process gene ontology terms as determined using DAVID for genes regulated, and upregulated and downregulated in ESCs+pRNA.

See also Figure S6.

Cell Stem Cell
The Nucleolus Regulates ESC Chromatin



(legend on next page)

EXPERIMENTAL PROCEDURES

mESC Culture

One hundred twenty-nine mouse embryonic stem cells (E14 line) were cultured in N2B27 media (Dulbecco's modified Eagle's medium [DMEM] F12, neurobasal medium, N2/B27 supplements, 2 mM L-glutamine with Pen/Strep, β-Mercaptoethanol) supplemented with recombinant leukemia inhibitory factor, LIF (ESGRO, 1000 U/ml) and MEK and GSK3β inhibitors, 2i (Stemolecule CHIR99021 and PD0325901, 3 μM and 1 μM, respectively). Cells were seeded at a density of 50,000 cells/cm^2 in culture dishes (Corning CellBIND surface) treated with 0.1% gelatine without feeder layer. Propagation of cells was carried out every 2 days using Trypsin 0.5× for enzymatic cell dissociation.

mESC Differentiation

mESCs were differentiated to neural progenitor cells according to a previously established protocol (Bibel et al., 2004). In brief, differentiation used a suspension-based embryoid bodies formation (Bacteriological Petri Dishes, Bio-one with vents, Greiner). The neural differentiation media (DMEM, 10% fetal calf serum, 1× MEM NEAA, 2 mM L-glutamine with Pen/Strep, β-mercaptoethanol) was filtered through 0.22 μm filters and stored at 4°C. During the 8-day differentiation procedure, media was exchanged every 2 days. In the last 4 days of differentiation, the media was supplemented with 2 μM retinoic acid to generate neural precursors that are Pax-6-positive radial glial cells.

Immunocytochemistry

ESCs or differentiated cells were grown on gelatin-coated coverslips and permeabilized with 0.05% Triton X-100 in 20 mMTris-HCl (pH 8), 5 mM MgCl2, 0.5 mM EDTA, and 25% glycerol. After washing, cells were fixed with cold methanol (7 min) and stained with anti-TIP5 (Diagenode) and anti-UBF (SantaCruz) antibodies and DAPI (Molecular Probes), and immunofluorescent images were digitally recorded.

Transmission Electron Microscopy

Cells were first fixed with 2.5% glutaraldehyde, dehydrated in an ethanol series, transferred to methanol, and immersed into a freshly prepared mixture of methanol and acetic anhydride (5:1, v/v) at 25°C for 24 hr in the dark (Tandler and Solari, 1982; Testillano et al., 1991). Cells were then washed in pure methanol for 20 min, transferred in ethanol and embedded in Epon (Sigma). Ultrathin (50 nm) sections were contrasted with 5% aqueous uranyl acetate for 60 min at room temperature and examined with a CM100 transmission electron microscope (FEI).

The volume of nucleolus-associated heterochromatin was estimated using the Cavalieri-estimator (Gundersen et al., 1988; West, 2012). Volume estimates were performed for samples of 25 to 34 nucleoli selected at random from each electron microscopy sample of two independent experiments. Nucleoli were selected independent of their size or shape in the electron microscopy montages. The volume of nucleolus-associated heterochromatin was expressed as a percentage of the volume of the corresponding nucleolus. Due to unequal variances of heterochromatin volumes in control and experimental cells, a heteroscedastic two-tailed t test was used to compare the groups ($p = 7.6 \times 10^{-9}$).

EMSA

Radiolabeled MCS-RNA was synthesized by T7 RNA polymerase using pBluescript-KS(+)/EcoRI as template. After treatment with DNase I, transcripts were purified and 50,000 cpm of MCS-RNA were incubated for 15 min on ice with 40 ng recombinant TIP5 or TTF1 in EMSA buffer (20 mM Tris-HCl [pH 8.0], 5 mM MgCl2, 100 mM KCl, and 0.2 mM EDTA). Cold competitor RNA was added, and incubation was continued for 30 min. RNA-protein complexes were analyzed by electrophoresis on 6% (w/v) native polyacrylamide gels and depicted with autoradiography.

For detailed methods, list of antibodies, and tables of primers, see the Supplemental Experimental Procedures.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, seven figures, and one table and can be found with this article online at http://dx.doi.org/10.1016/j.stem.2014.10.005.

ACKNOWLEDGMENTS

The authors thank Ursula Luethi, Lutz Slomianka, and the Center for Microscopy and Image Analysis, University of Zurich for performing transmission electron microscopy analyses. This work was supported by the Swiss National Science Foundation (310003A-135801 and 31003A-152854 to R.S., 31003A-118361 to P.C., and 323530-133905 to F.A.W.), UBS-Promedica Stiftung (to R.S. and O.S.), Forschungskredit of the University of Zurich (to N.S. and E.V.), and the Mäxi Stiftung (to D.B. and R.S.).

Received: February 27, 2014
Revised: August 8, 2014
Accepted: October 16, 2014
Published: December 4, 2014

REFERENCES

Aoto, T., Saitoh, N., Ichimura, T., Niwa, H., and Nakao, M. (2006). Nuclear and chromatin reorganization in the MHC-Oct3/4 locus at developmental phases of embryonic stem cell differentiation. Dev. Biol. *298*, 354–367.

Bártová, E., Galiová, G., Krejcí, J., Harnicarová, A., Strasák, L., and Kozubek, S. (2008a). Epigenome and chromatin structure in human embryonic stem cells undergoing differentiation. Dev. Dyn. *237*, 3690–3702.

Bártová, E., Krejcí, J., Harnicarová, A., and Kozubek, S. (2008b). Differentiation of human embryonic stem cells induces condensation of chromosome territories and formation of heterochromatin protein 1 foci. Differentiation *76*, 24–32.

Bhattacharya, D., Talwar, S., Mazumder, A., and Shivashankar, G.V. (2009). Spatio-temporal plasticity in chromatin organization in mouse cell differentiation and during Drosophila embryogenesis. Biophys. J. *96*, 3832–3839.

Bibel, M., Richter, J., Schrenk, K., Tucker, K.L., Staiger, V., Korte, M., Goetz, M., and Barde, Y.A. (2004). Differentiation of mouse embryonic stem cells into a defined neuronal lineage. Nat. Neurosci. *7*, 1003–1009.

Figure 7. Depletion of TIP5 Impairs ESC Differentiation

(A) Schema showing the experimental strategy for TIP5 knockdown in ESCs.
(B) *Tip5*, *Oct4*, *Nanog*, and *Rex1* mRNA levels in ESCs depleted of TIP5.
(C) TIP5 knockdown affects ESC proliferation. Data represent relative cell numbers after 3 days of siRNA treatment.
(D and E) (D) Cell morphology and (E) AP staining of ESCs treated with siRNA-control and -*Tip5*.
(F–H) (F) Cell morphology and AP staining of cells after 3 days of differentiation. Quantifications are shown respectively in (G) differentiated (+), partially differentiated (+/−), and not differentiated (−); and (H) stained (+), partially stained (+/−), and not stained (−). See also Figure S7.
(I) Model showing the chromatin organization of the nucleus and nucleolus of ESCs (open, euchromatic) and differentiated cells (closed, heterochromatic). In ESCs, IGS-rRNA is not processed with consequent lack of mature pRNA. The unprocessed transcript impairs the association of TIP5 with TTF1, inhibiting TIP5 recruitment to rDNA and causing the euchromatic state of all rRNA genes. Upon differentiation, mature pRNA is produced and allows TIP5-TTF1 interaction that is productive for TIP5 targeting to rDNA and formation of heterochromatin at nucleoli. The arrow depicts the influence of rDNA heterochromatin in the formation of highly condensed and repressive chromatin structures at region outside the nucleolus. Establishment of compact heterochromatic structures abrogates genome plasticity of ESCs and set up a genome organization that favors cell lineage specification.

Bierhoff, H., Postepska-Igielska, A., and Grummt, I. (2013). Noisy silence: Non-coding RNA and heterochromatin formation at repetitive elements. Epigenetics *9*, 53–61.

Cammas, L., Romand, R., Fraulob, V., Mura, C., and Dolle, P. (2007). Expression of the murine retinol dehydrogenase 10 (Rdh10) gene correlates with many sites of retinoid signalling during embryogenesis and organ differentiation. Developmental dynamics *236*, 2899–2908.

Carvalho, C., Pereira, H.M., Ferreira, J., Pina, C., Mendonça, D., Rosa, A.C., and Carmo-Fonseca, M. (2001). Chromosomal G-dark bands determine the spatial organization of centromeric heterochromatin in the nucleus. Mol. Biol. Cell *12*, 3563–3572.

Conconi, A., Widmer, R.M., Koller, T., and Sogo, J.M. (1989). Two different chromatin structures coexist in ribosomal RNA genes throughout the cell cycle. Cell *57*, 753–761.

de Wit, E., Bouwman, B.A., Zhu, Y., Klous, P., Splinter, E., Verstegen, M.J., Krijger, P.H., Festuccia, N., Nora, E.P., Welling, M., et al. (2013). The pluripotent genome in three dimensions is shaped around pluripotency factors. Nature *501*, 227–231.

Dev, V.G., Tantravahi, R., Miller, D.A., and Miller, O.J. (1977). Nucleolus organizers in Mus musculus subspecies and in the RAG mouse cell line. Genetics *86*, 389–398.

Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature *485*, 376–380.

Efroni, S., Duttagupta, R., Cheng, J., Dehghani, H., Hoeppner, D.J., Dash, C., Bazett-Jones, D.P., Le Grice, S., McKay, R.D., Buetow, K.H., et al. (2008). Global transcription in pluripotent embryonic stem cells. Cell Stem Cell *2*, 437–447.

Evers, R., and Grummt, I. (1995). Molecular coevolution of mammalian ribosomal gene terminator sequences and the transcription termination factor TTF-I. Proc. Natl. Acad. Sci. USA *92*, 5827–5831.

Fussner, E., Ahmed, K., Dehghani, H., Strauss, M., and Bazett-Jones, D.P. (2010). Changes in chromatin fiber density as a marker for pluripotency. Cold Spring Harb. Symp. Quant. Biol. *75*, 245–249.

Gaspar-Maia, A., Alajem, A., Meshorer, E., and Ramalho-Santos, M. (2011). Open chromatin in pluripotency and reprogramming. Nat. Rev. Mol. Cell Biol. *12*, 36–47.

Gerber, J.K., Gögel, E., Berger, C., Wallisch, M., Müller, F., Grummt, I., and Grummt, F. (1997). Termination of mammalian rDNA replication: polar arrest of replication fork movement by transcription termination factor TTF-I. Cell *90*, 559–567.

Gonzalez-Sandoval, A., Towbin, B.D., and Gasser, S.M. (2013). The formation and sequestration of heterochromatin during development: delivered on 7 September 2012 at the 37th FEBS Congress in Sevilla, Spain. FEBS J. *280*, 3212–3219.

Gorkin, D.U., Leung, D., and Ren, B. (2014). The 3D genome in transcriptional regulation and pluripotency. Cell Stem Cell *14*, 762–775.

Guetg, C., Lienemann, P., Sirri, V., Grummt, I., Hernandez-Verdun, D., Hottiger, M.O., Fussenegger, M., and Santoro, R. (2010). The NoRC complex mediates the heterochromatin formation and stability of silent rRNA genes and centromeric repeats. EMBO J. *29*, 2135–2146.

Guetg, C., Scheifele, F., Rosenthal, F., Hottiger, M.O., and Santoro, R. (2012). Inheritance of silent rDNA chromatin is mediated by PARP1 via noncoding RNA. Mol. Cell *45*, 790–800.

Gundersen, H.J.G., Bendtsen, T.F., Korbo, L., Marcussen, N., Møller, A., Nielsen, K., Nyengaard, J.R., Pakkenberg, B., Sørensen, F.B., Vesterby, A., et al. (1988). Some new, simple and efficient stereological methods and their use in pathological research and diagnosis. APMIS *96*, 379–394.

Hawkins, R.D., Hon, G.C., Lee, L.K., Ngo, Q., Lister, R., Pelizzola, M., Edsall, L.E., Kuan, S., Luu, Y., Klugman, S., et al. (2010). Distinct epigenomic landscapes of pluripotent and lineage-committed human cells. Cell Stem Cell *6*, 479–491.

Huang, W., Sherman, B.T., and Lempicki, R.A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. *4*, 44–57.

Jørgensen, H.F., Azuara, V., Amoils, S., Spivakov, M., Terry, A., Nesterova, T., Cobb, B.S., Ramsahoye, B., Merkenschlager, M., and Fisher, A.G. (2007). The impact of chromatin modifiers on the timing of locus replication in mouse embryonic stem cells. Genome Biol. *8*, R169.

Kind, J., Pagie, L., Ortabozkoyun, H., Boyle, S., de Vries, S.S., Janssen, H., Amendola, M., Nolen, L.D., Bickmore, W.A., and van Steensel, B. (2013). Single-cell dynamics of genome-nuclear lamina interactions. Cell *153*, 178–192.

Kurihara, Y., Suh, D.S., Suzuki, H., and Moriwaki, K. (1994). Chromosomal locations of Ag-NORs and clusters of ribosomal DNA in laboratory strains of mice. Mamm. Genome *5*, 225–228.

Längst, G., Blank, T.A., Becker, P.B., and Grummt, I. (1997). RNA polymerase I transcription on nucleosomal templates: the transcription termination factor TTF-I induces chromatin remodeling and relieves transcriptional repression. EMBO J. *16*, 760–768.

Leitch, H.G., McEwen, K.R., Turp, A., Encheva, V., Carroll, T., Grabole, N., Mansfield, W., Nashun, B., Knezovich, J.G., Smith, A., et al. (2013). Naive pluripotency is associated with global DNA hypomethylation. Nat. Struct. Mol. Biol. *20*, 311–316.

Martens, J.H., O'Sullivan, R.J., Braunschweig, U., Opravil, S., Radolf, M., Steinlein, P., and Jenuwein, T. (2005). The profile of repeat-associated histone lysine methylation states in the mouse epigenome. EMBO J. *24*, 800–812.

Mayer, C., Schmitz, K.M., Li, J., Grummt, I., and Santoro, R. (2006). Intergenic transcripts regulate the epigenetic state of rRNA genes. Mol. Cell *22*, 351–361.

Mayer, C., Neubert, M., and Grummt, I. (2008). The structure of NoRC-associated RNA is crucial for targeting the chromatin remodelling complex NoRC to the nucleolus. EMBO Rep. *9*, 774–780.

Melcer, S., Hezroni, H., Rand, E., Nissim-Rafinia, M., Skoultchi, A., Stewart, C.L., Bustin, M., and Meshorer, E. (2012). Histone modifications and lamin A regulate chromatin protein dynamics in early embryonic stem cell differentiation. Nat. Commun. *3*, 910.

Meshorer, E., and Misteli, T. (2006). Chromatin in pluripotent embryonic stem cells and differentiation. Nat. Rev. Mol. Cell Biol. *7*, 540–546.

Meshorer, E., Yellajoshula, D., George, E., Scambler, P.J., Brown, D.T., and Misteli, T. (2006). Hyperdynamic plasticity of chromatin proteins in pluripotent embryonic stem cells. Dev. Cell *10*, 105–116.

Moss, T., and Stefanovsky, V.Y. (2002). At the center of eukaryotic life. Cell *109*, 545–548.

Németh, A., Strohner, R., Grummt, I., and Längst, G. (2004). The chromatin remodeling complex NoRC and TTF-I cooperate in the regulation of the mammalian rRNA genes in vivo. Nucleic Acids Res. *32*, 4091–4099.

Nora, E.P., Lajoie, B.R., Schulz, E.G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., van Berkum, N.L., Meisig, J., Sedat, J., et al. (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature *485*, 381–385.

Paredes, S., and Maggert, K.A. (2009). Ribosomal DNA contributes to global chromatin regulation. Proc. Natl. Acad. Sci. USA *106*, 17829–17834.

Pinheiro, I., Margueron, R., Shukeir, N., Eisold, M., Fritzsch, C., Richter, F.M., Mittler, G., Genoud, C., Goyama, S., Kurokawa, M., et al. (2012). Prdm3 and Prdm16 are H3K9me1 methyltransferases required for mammalian heterochromatin integrity. Cell *150*, 948–960.

Rinn, J.L., and Chang, H.Y. (2012). Genome regulation by long noncoding RNAs. Annu. Rev. Biochem. *81*, 145–166.

Santoro, R., and Grummt, I. (2001). Molecular mechanisms mediating methylation-dependent silencing of ribosomal gene transcription. Mol. Cell *8*, 719–725.

Santoro, R., and Grummt, I. (2005). Epigenetic mechanism of rRNA gene silencing: temporal order of NoRC-mediated histone modification, chromatin remodeling, and DNA methylation. Mol. Cell. Biol. *25*, 2539–2546.

Santoro, R., Li, J., and Grummt, I. (2002). The nucleolar remodeling complex NoRC mediates heterochromatin formation and silencing of ribosomal gene transcription. Nat. Genet. *32*, 393–396.

Santoro, R., Schmitz, K.M., Sandoval, J., and Grummt, I. (2010). Intergenic transcripts originating from a subclass of ribosomal DNA repeats silence ribosomal RNA genes in trans. EMBO Rep. *11*, 52–58.

Schlesinger, S., Selig, S., Bergman, Y., and Cedar, H. (2009). Allelic inactivation of rDNA loci. Genes Dev. *23*, 2437–2447.

Schmitz, K.M., Mayer, C., Postepska, A., and Grummt, I. (2010). Interaction of noncoding RNA with the rDNA promoter mediates recruitment of DNMT3b and silencing of rRNA genes. Genes Dev. *24*, 2264–2269.

Splinter, E., and de Laat, W. (2011). The complex transcription regulatory landscape of our genome: control in three dimensions. EMBO J. *30*, 4345–4355.

Strohner, R., Nemeth, A., Jansa, P., Hofmann-Rohrer, U., Santoro, R., Längst, G., and Grummt, I. (2001). NoRC—a novel member of mammalian ISWI-containing chromatin remodeling machines. EMBO J. *20*, 4892–4900.

Tandler, C.J., and Solari, A.J. (1982). Methanol-acetic anhydride: an efficient blocking agent for electron microscope cytochemistry. Its application to mouse testis and other tissues. Histochemistry *76*, 351–361.

Testillano, P.S., Sanchez-Pina, M.A., Olmedilla, A., Ollacarizqueta, M.A., Tandler, C.J., and Risueno, M.C. (1991). A specific ultrastructural method to reveal DNA: the NAMA-Ur. The journal of histochemistry and cytochemistry: official journal of the Histochemistry Society *39*, 1427–1438.

Towbin, B.D., González-Aguilera, C., Sack, R., Gaidatzis, D., Kalck, V., Meister, P., Askjaer, P., and Gasser, S.M. (2012). Step-wise methylation of histone H3K9 positions heterochromatin at the nuclear periphery. Cell *150*, 934–947.

Wen, B., Wu, H., Shinkai, Y., Irizarry, R.A., and Feinberg, A.P. (2009). Large histone H3 lysine 9 dimethylated chromatin blocks distinguish differentiated from embryonic stem cells. Nat. Genet. *41*, 246–250.

West, M.J. (2012). Estimating volume in biological structures. Cold Spring Harb Protoc. *2012*, 1129–1139.

Wong, L.H., Ren, H., Williams, E., McGhie, J., Ahn, S., Sim, M., Tam, A., Earle, E., Anderson, M.A., Mann, J., and Choo, K.H. (2009). Histone H3.3 incorporation provides a unique and functionally essential telomeric chromatin in embryonic stem cells. Genome Res. *19*, 404–414.

Yoshida, Y., Matsuda, S., Ikematsu, N., Kawamura-Tsuzuku, J., Inazawa, J., Umemori, H., and Yamamoto, T. (1998). ANA, a novel member of Tob/BTG1 family, is expressed in the ventricular zone of the developing central nervous system. Oncogene *16*, 2687–2693.

Zhou, Y., Santoro, R., and Grummt, I. (2002). The chromatin remodeling complex NoRC targets HDAC1 to the ribosomal gene promoter and represses RNA polymerase I transcription. EMBO J. *21*, 4632–4640.

Cell Stem Cell, Volume *15*

Supplemental Information

lncRNA Maturation to Initiate Heterochromatin Formation in the Nucleolus Is Required for Exit from Pluripotency in ESCs

Nataša Savić, Dominik Bär, Sergio Leone, Sandra C. Frommel, Fabienne A. Weber, Eva Vollenweider, Elena Ferrari, Urs Ziegler, Andres Kaech, Olga Shakhova, Paolo Cinelli, and Raffaella Santoro

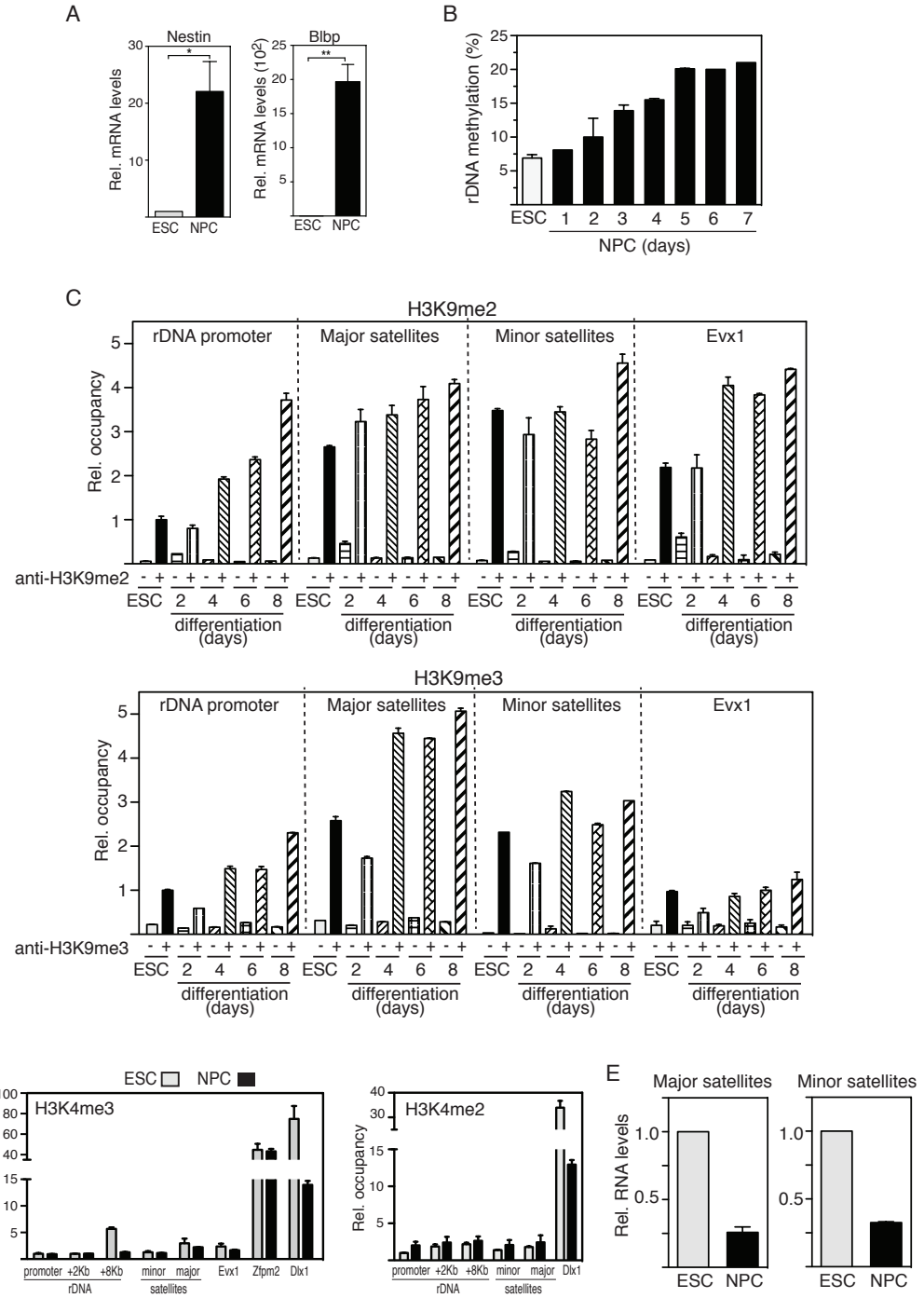

A
Nestin
Blbp
Rel. mRNA levels
Rel. mRNA levels (10^2)
ESC NPC
B
rDNA methylation (%)
ESC 1 2 3 4 5 6 7
NPC (days)
C
H3K9me2
rDNA promoter
Major satellites
Minor satellites
Evx1
Rel. occupancy
anti-H3K9me2
ESC 2 4 6 8
differentiation (days)
H3K9me3
anti-H3K9me3
D
ESC NPC
H3K4me3
H3K4me2
promoter +2Kb +8Kb minor major Evx1 Zfpm2 Dlx1
rDNA
satellites
E
Major satellites
Minor satellites
Rel. RNA levels
ESC NPC

Figure S1

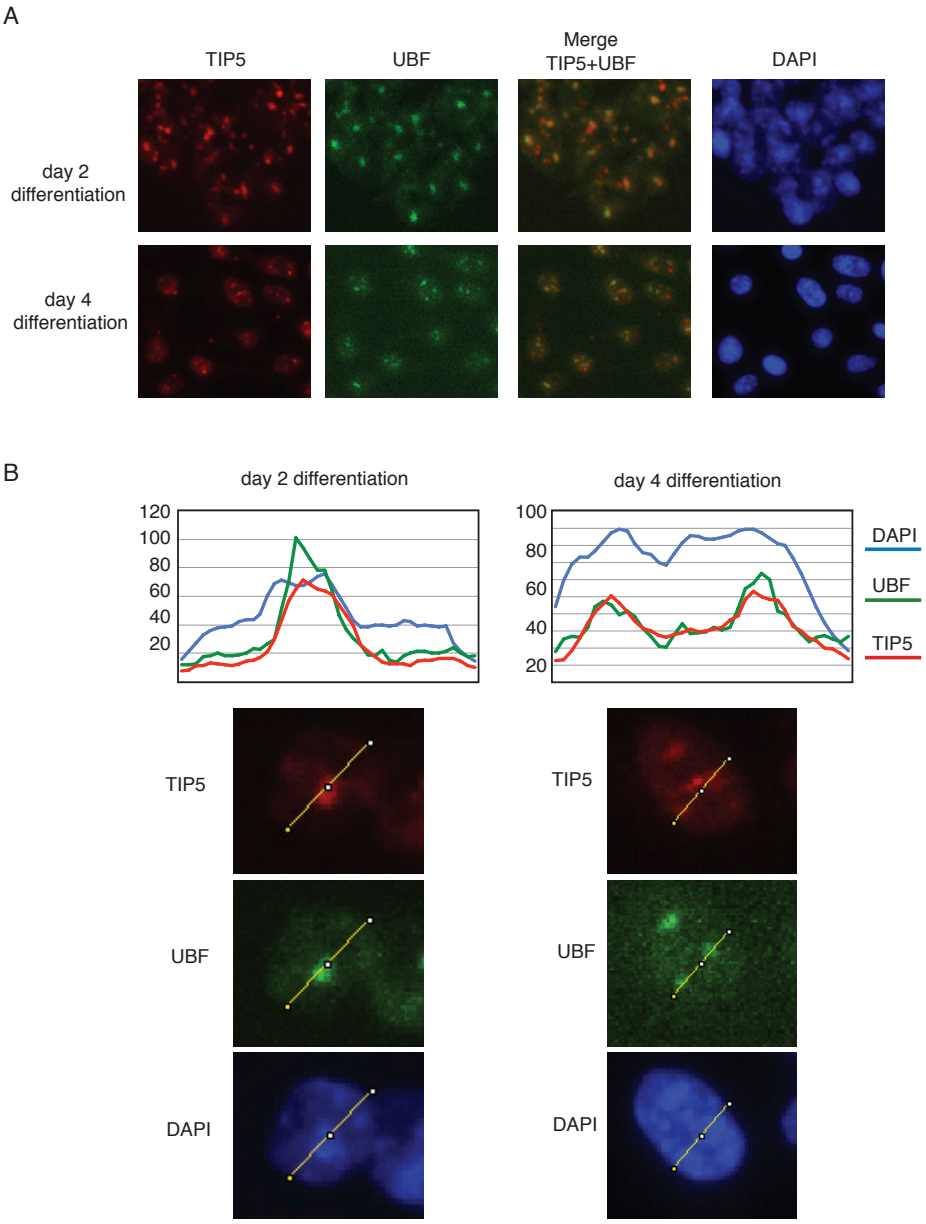

A
TIP5
UBF
Merge
TIP5+UBF
DAPI
day 2
differentiation
day 4
differentiation
B
day 2 differentiation
day 4 differentiation
120
100
80
60
40
20
100
80
60
40
20
DAPI
UBF
TIP5
TIP5
UBF
DAPI
TIP5
UBF
DAPI

Figure S2

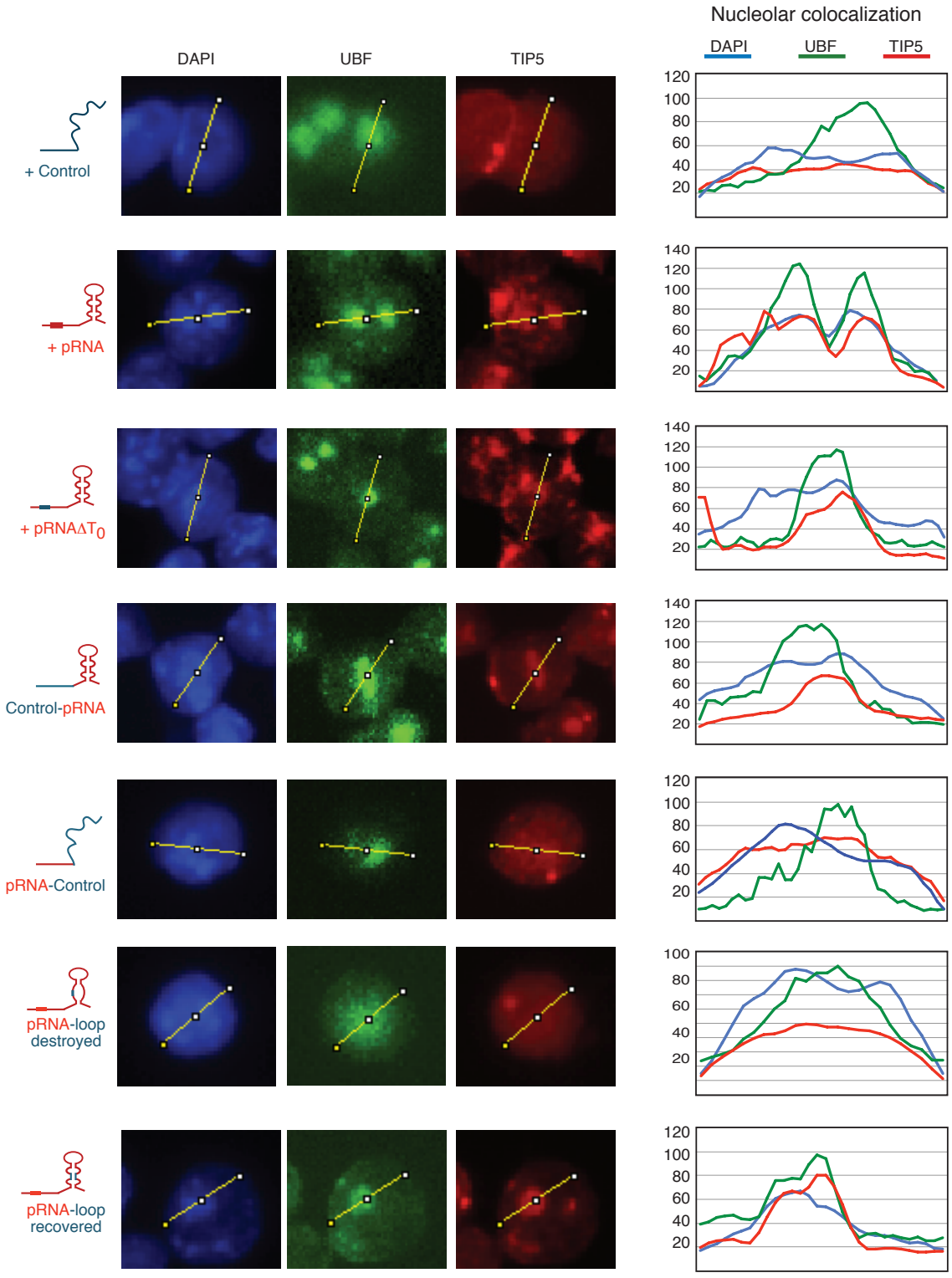

DAPI
UBF
TIP5
Nucleolar colocalization
DAPI
UBF
TIP5
+ Control
+ pRNA
+ pRNAΔT0
Control-pRNA
pRNA-Control
pRNA-loop destroyed
pRNA-loop recovered

Figure S3

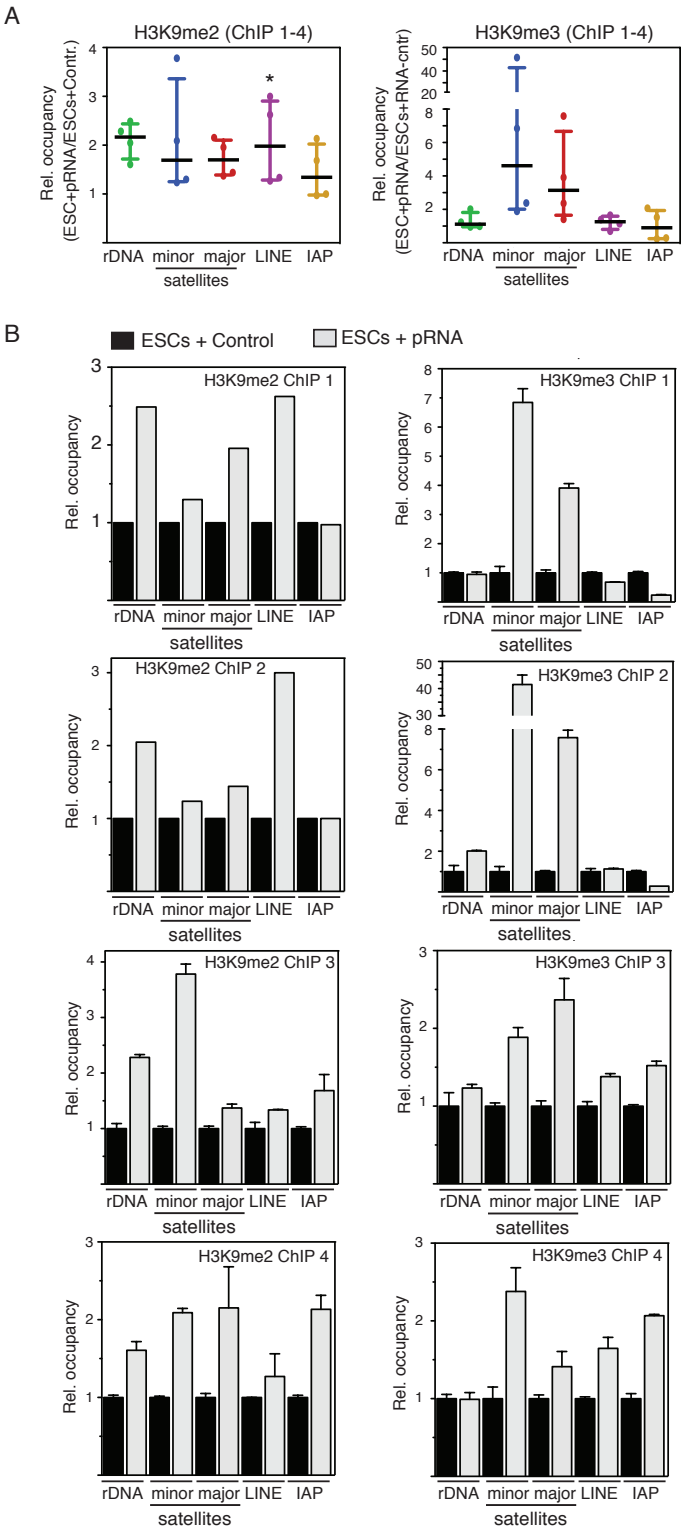

A
H3K9me2 (ChIP 1-4)
H3K9me3 (ChIP 1-4)
Rel. occupancy (ESC+pRNA/ESCs+Contr.)
Rel. occupancy (ESC+pRNA/ESCs+RNA-cntr)
rDNA minor major LINE IAP
satellites
B
ESCs + Control
ESCs + pRNA
H3K9me2 ChIP 1
H3K9me3 ChIP 1
H3K9me2 ChIP 2
H3K9me3 ChIP 2
H3K9me2 ChIP 3
H3K9me3 ChIP 3
H3K9me2 ChIP 4
H3K9me3 ChIP 4
Rel. occupancy

Figure S4

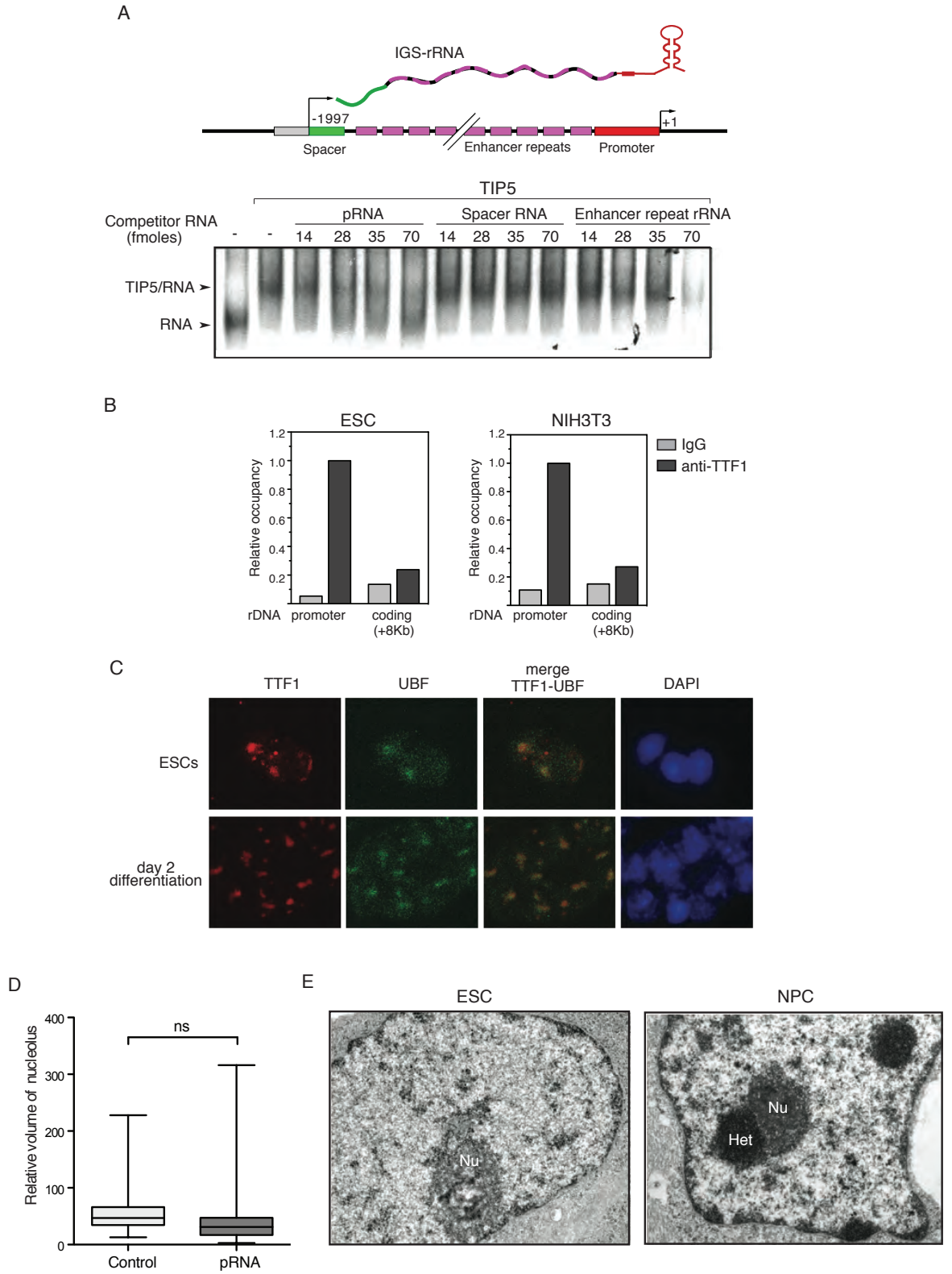

A
IGS-rRNA
-1997
+1
Spacer
Enhancer repeats
Promoter
TIP5
Competitor RNA (fmoles)
pRNA
Spacer RNA
Enhancer repeat rRNA
- - 14 28 35 70 14 28 35 70 14 28 35 70
TIP5/RNA
RNA
B
ESC
NIH3T3
IgG
anti-TTF1
Relative occupancy
1.2
1.0
0.8
0.6
0.4
0.2
rDNA promoter coding (+8Kb)
C
TTF1
UBF
merge TTF1-UBF
DAPI
ESCs
day 2 differentiation
D
Relative volume of nucleolus
400
300
200
100
0
ns
Control
pRNA
E
ESC
NPC
Nu
Het

Figure S5

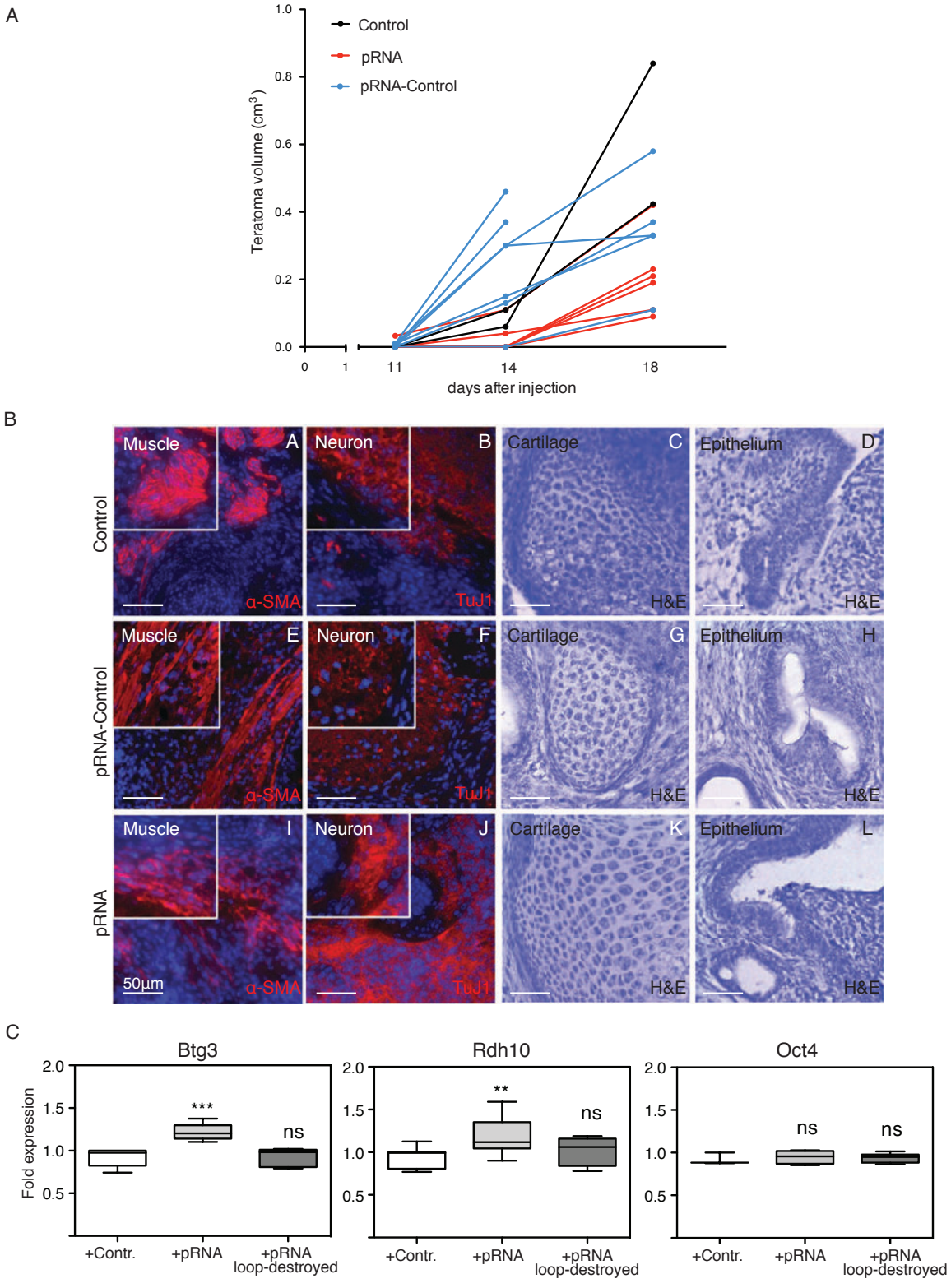

A
Control
pRNA
pRNA-Control
Teratoma volume (cm3)
days after injection
B
Control
pRNA-Control
pRNA
Muscle
Neuron
Cartilage
Epithelium
α-SMA
TuJ1
H&E
50µm
C
Btg3
Rdh10
Oct4
Fold expression
+Contr.
+pRNA
+pRNA loop-destroyed
***
**
ns

Figure S6

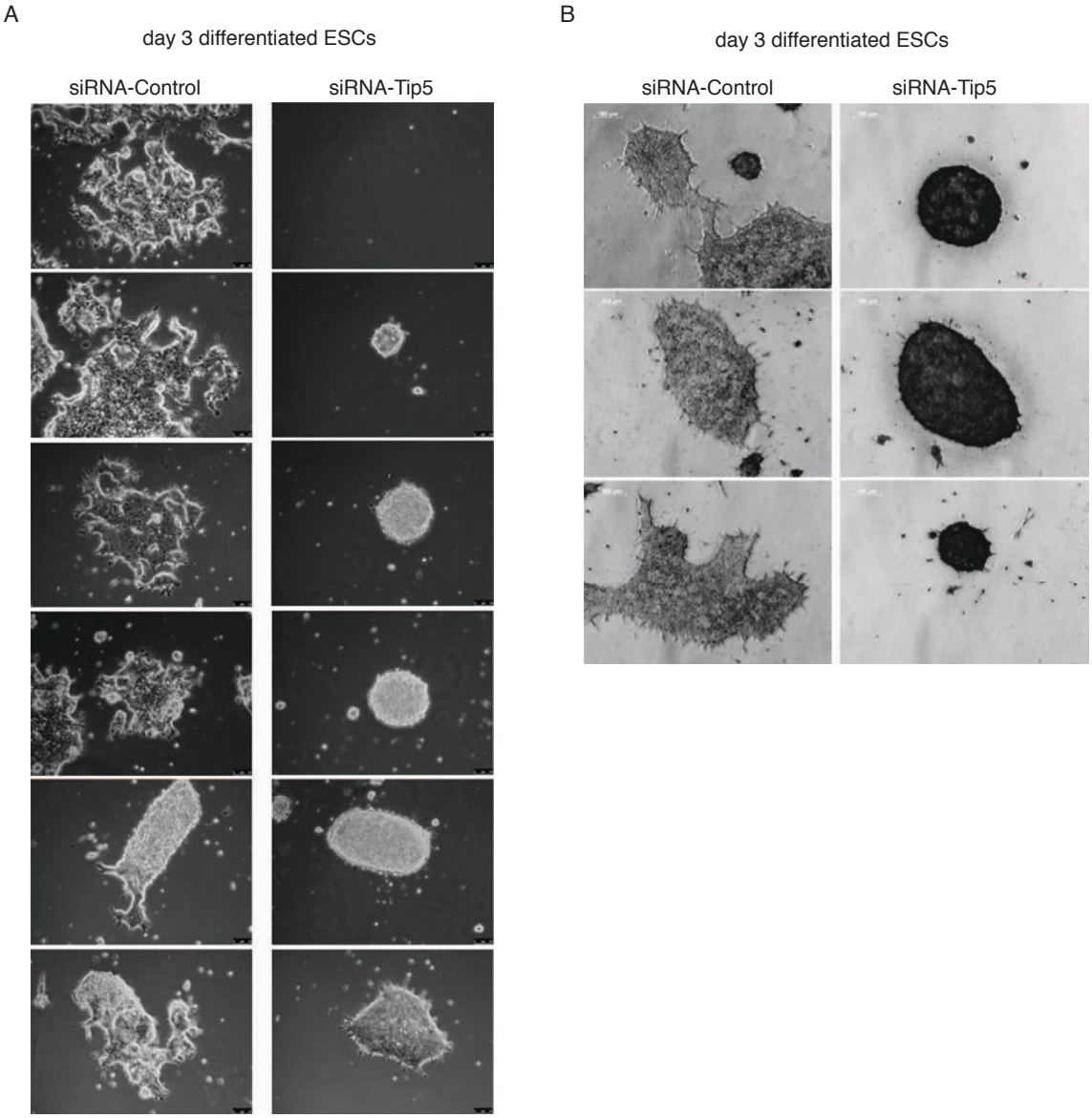

A
day 3 differentiated ESCs
siRNA-Control
siRNA-Tip5
B
day 3 differentiated ESCs
siRNA-Control
siRNA-Tip5

Figure S7

Supplemental Figure Legends

Figure S1. Establishment of rDNA heterochromatin occurs during ESC differentiation. Related to Figure 1.

(A) qRT-PCR. *Nestin* and *Blbp* (neural precursor markers) mRNA levels in ESCs and NPCs. Data were normalized to *Rps12* mRNA. Error bars indicate the SD of three independent experiments.

(B) CpG methylation levels at rDNA promoter in ESCs (JM8N4) and during 7 days of differentiation into NPCs. ESCs were cultivated on monolayer and differentiation was induced with N2B27 medium supplemented with RA. Error bars indicate the SD of two independent experiments.

(C) ChIP. H3K9me2 and H3K9me3 occupancy at rDNA promoter and coding sequences, major and minor satellites and control gene *Evx1* monitored during different time points (days) of differentiation. Data were normalized to input and rDNA promoter values in ESCs.

(D) ChIP. H3K4me2 and H3K4me3 occupancy in ESCs and NPCs. *Evx1*, *Zfpm2* and *Dlx1* represent control genes. Data were normalized to input and rDNA promoter values in ESCs.

(E) Major and minor satellite transcript levels in ESCs and NPCs were measured by qRT-PCR and normalized to *Rps12* mRNA. Error bars indicate the SD of two independent experiments.

Figure S2. Establishment of rDNA heterochromatin during ESC differentiation correlates with the recruitment of TIP5 to rDNA. Related to Figure 1.

(A) TIP5 localizes within nucleoli shortly after ESC differentiation. Immunofluorescence showing TIP5 nucleolar localization in ESCs after 2 and 4 days differentiation. Nucleoli are visualized by UBF signal.

(B) Quantification of TIP5, UBF and DAPI colocalization measured using Fiji image analysis software.

Figure S3. The stem-loop structure of pRNA is sufficient to target TIP5 to nucleoli. Related to Figure 3.

Quantification of nucleolar localization of TIP5 in ESCs transfected with RNA Control, pRNA, pRNA and pRNA mutants (pRNAΔT_0, Control-pRNA, pRNA-Control, pRNA-loop destroyed and pRNA-loop recovered). TIP5, UBF and DAPI signals were measured using Fiji image analysis software.

Figure S4. Mature pRNA increases H3K9me2 and H3K9me3 at rDNA and centric-pericentric heterochromatin in ESCs. Related to Figure 3.

(A) Scatter plot of the four ChIP experiments shown in Figure 3D.

(B) Results of the single four ChIP experiments showing an inverse correlation in the enrichment between H3K9me2 and H3K9me3 levels at minor and major repeats in ESCs+pRNA when compared to ESCs+RNA-control. Data are represented as bound over input in ESCs+pRNA normalized to values measured in ESCs+RNA-control.

Figure S5. Mature pRNA mediates TIP5-TTF1 interaction and induces global remodelling toward heterochromatic structures pRNA. Related to Figure 4 and 5.

(A) TIP5 binds to IGS-rRNA and it has a stronger affinity for pRNA sequences. Increasing equal moles of *in vitro* transcripts corresponding to pRNA, spacer promoter and enhancer repeat RNA were used to compete for binding of $TIP5_{332-723}$ to radiolabelled run-off transcripts from pBluescript (MCS-RNA). RNA/protein complexes were analyzed by EMSA.

(B) TTF1 binds to the rDNA promoter of ESCs. ChIP showing association of TTF1 with rDNA promoter in ESCs and NIH3T3 cells. Data of two independent experiments were normalized to input and rDNA promoter values. The low levels of TTF1 association with +8 Kb rDNA sequences (that do not contain T elements) demonstrated the specificity of the assay.

(C) TTF1 is localized within nucleoli of ESCs and differentiated cells. Immunofluorescence showing TTF1 nucleolar localization in ESCs and 2 days after differentiation. Nucleoli are visualized by UBF signal.

(D) Box-and-whisker plot of nucleoli size in ESCs+Control and ESCs+pRNA. 25 to 34 nucleoli of ESCs+RNA control and ESCs+pRNA of two independent experiments were selected at random

and independent of their size or shape in the EM montages. Volumes were estimated using the Cavalieri-estimator (Gundersen et al., 1988; West, 2012). Nucleoli volumes did not differ among control and ESCs+pRNA while heterochromatin associated to nucleoli did. The volume of nucleolus-associated heterochromatin was expressed as a percentage of the volume of the nucleolus (Nu) that was associated with and shown in main Figure 5B.

(E) Transmission electron microscopy analysis of ESCs and NPCs. The contrast procedure reveals in dark condensed heterochromatic structures (Het) and nucleoli (Nu).

Figure S6. pRNA impairs ESC pluripotency. Related to Figure 6.

(A) Kinetics of teratoma growth from the time of injection to euthanasia of control animals. In this experiment, tumors derived from ESCs+RNA-control (2), ESCs+pRNA (6) and ESCs+mutant pRNA-Control (7) were analysed.

(B) Histological analysis of teratomas derived from ESCs transfected with RNA-control, pRNA and mutant pRNA-Control revealed that ESCs differentiate into all three germ layers as shown by the presence of ectoderm (B, F, J), endoderm (C) and mesoderm (A,E,I). C,D,G,H,K,L haematoxylin staining. Immunostaining for bIII tubulin (TuJ1) (B,F,J), and smooth muscle actin (SMA) (A,E,I). Inserts show higher magnification. Scale bars, 50 μm.

(C) Box-and-whisker plot of three independent experiments showing mRNA levels of *Btg3, Rdh10* and *Oct4* measured in ESCs transfected with RNA-control, pRNA and mutant pRNA loop-destroyed. *Btg3* and *Rdh10* are known to be implicated in neurogenesis and embryonic differentiation (Cammas et al., 2007; Yoshida et al., 1998). Consistent with the results of Fig. 6A, *Oct4* mRNA levels remained unaffected in all three conditions.

Figure S7. Depletion of TIP5 impairs ESC differentiation. Related to Figure 7.

(A) Cell morphology and (B) AP staining of differentiated ESCs treated with siRNA-control and -*Tip5*.

Supplemental Table Legends

Supplemental Table S1. Related to Figure 6.

Total RNA of ESCs+RNA-Control and ESCs+pRNA from two biological replicates were purified and analyzed by RNA seq. The table includes the list of genes whose transcript levels were altered in ESCs+pRNA when compared to control cells (defined as regulated, upregulated and downregulated) and gene onthology analysis using DAVID tools.

Supplemental Experimental procedures

Reprogramming into iPSC

Reprogramming was performed as previously described (Weber et al., 2013). Breifly, mouse embryonic fibroblasts (MEF) were isolated from 14.5 day-pregnant C57BL/6 mice and cultured in DMEM supplemented with 10% FBS (PAA) and 1% L-glutamin/penicillin/streptomycin (10,000 U/ml penicillin G sodium; 10,000µg /ml streptomycin sulphate; 29.2mg/ml L-glutamine; 10mM sodium citrate in 0.14% NaCl, Gibco). The reprogramming of the MEFs was performed according to Yamanaka's protocol (Takahashi et al., 2007) using the pMXs retroviral vectors producing murine *Oct4*, *Sox2*, *Klf4* and *c-Myc* (Addgene, cat. nos. 13366, 13367, 13370 and 13375). Two days after infection, MEFs were cultured in DMEM containing 15% FBS, 1% L-glutamin/penicillin/streptomycin, 1x MEM non-essential amino acids (GIBCO) and 50 mM ß-mercaptoethanol (GIBCO) supplemented with 1000 U/ml ESGRO murine Leukemia inhibitory factor (LIF, Chemicon Int.). The iPSC cell line used for the experiment of Figure 1D has the ability to generate teratoma (data not shown).

Transfections

ESCs were seeded at a density of 20,000 cells/cm^2 and transfected with the indicated siRNAs (50 nM siRNA) or synthetic RNAs (1 mg/ml) using Lipofectamine® RNAiMAX (Life Techonlogies) in Opti-MEM® GlutaMAX™ (GIBCO) reduced-serum medium. Analysis of differentiated transfected ESCs was performed using consecutive transfections. Three days after the first transfection, equal amounts of ESCs (e.g. siRNA-control and siTIP5 treated cells) were again transfected and induced to differentiate in complete media (G-MEM, 10%FCS, Sodium Pyruvate 100mM, 1xMEM NEAA, L-Glutamine) by withdrawal of LIF and 2i. Efficiencies of siRNA-mediated depletions and synthetic RNA levels were monitored by qRT-PCR 3-4 days post-transfection.

In Vitro Transcription

The indicated pRNA and control sequences were cloned by PCR into pJET1/2 plasmids. pRNA : mrDNA from -232 to -1; Control-pRNA: control sequences at 5', mrDNA from -140 to -1 at 3';

pRNA-Control: mrDNA from -232 to -140 at 5', control sequences at 3' ; pRNA-loop destroyed: mrDNA from -232 to -1 sequences where GGG (-115/-113) were replaced with AAA; pRNA-loop recovered: mrDNA from -232 to -1 sequences where CCC (-60/-58) were replaced with TTT. All plasmids were verified by sequencing. Synthetic RNAs were synthesized using T7 polymerase and as substrate Xba I linearized pJET1/2 vectors containing the indicated sequences. After treatment with DNase I, transcripts were double purified using TRIzol reagent (Invitrogen) according to the manufacture's protocol.

Chromatin Immunoprecipitation

The chromatin immunoprecipitation (ChIP) protocol was previously described (Santoro, 2014). Briefly, formaldehyde 1% was added to cultured cells to cross-link proteins to DNA. Isolated nuclei were then lysed in 300µl lysis buffer (50 mM Tris-HCl [pH 8.1], 10 mM EDTA, 1% SDS) and sonicated using a Bioruptor ultrasonic cell disruptor to shear genomic DNA to an average fragment size of 200bp. 20 to 40 mg chromatin was diluted tenfold with IP buffer (16.7 mM Tris-HCl [pH 8.1]), 167 mM NaCl, 1.2 mM EDTA, 0.01% SDS, 1.1% Triton X-100) and than immunoprecipitated overnight with ChIP-grade antibodies. After elution and reversion of crosslinks, the precipitated DNA was purified with phenol/chloroform, ethanol precipitated and than quantified by qPCR. rDNA, major and minor satellite sequences were amplified with previously reported primers (Martens et al., 2005) (Martens et al., 2005; Santoro et al., 2002). Primers are listed in Table S2.

RNA Extraction, reverse transcription and quantitative PCR (RT-qPCR)

RNA was purified with Trizol reagent (Life Technologies) . Residual contaminating genomic DNA was removed with Ambion® TURBO™ DNase according to manufacture's instructions. RNA was primed with random hexamers and reverse-transcribed to first-strand cDNA. Reverse transcription of pRNA and IGS-rRNA was performed using DNA oligo -20/-1 Rev or random primers. qRT-PCR was performed with SensiMix SYBR Hi-ROX Mix (Bioline) on a Rotor-Gene Q (Qiagen). Amplification of samples without reverse transcriptase assured absence of DNA (data

not shown). The relative transcription levels were determined by normalizing to Rps12 mRNA levels. Statistical significance (P-values) of the difference in expression levels between genes was calculated using the two-sample paired t-test. Primer sequences used in qRT-PCR are listed in Table S2.

CpG methylation

rDNA CpG methylation was measured as previously described (Santoro, 2014; Santoro et al., 2002)). 2 μg genomic DNA were digested with HpaII (NEB) in the presence of 5 ng of unmethylated pBluescript KS(+) plasmid. rDNA CpG methylation levels were measured by quantitative amplification using primer pairs (-165/-145 Forw and -20/-1 Rev) that flank the restriction sites CCGG at -142 of rDNA promoter or primers that amplify neighbouring sequences lacking HpaII sites (+1/+20 Forw and +111/+130 Rev). Values were obtained using logarithmic dilutions of mouse genomic DNA as standard curve. CpG methylation levels were calculated as resistance to HpaII digestion by normalizing the amounts of rDNA amplified from -165 to -1 to the levels of amplicons from +1 to +130. To verify HpaII digestion efficiency, pBluescript KS(+) plasmid was analyzed by qPCR using one forward primer that is complementary to sequences upstream of the CCGG site of β-lactamase gene (at 2580) and two different reverse primers that map upstream and downstream the HpaII sites (see Table S2). All analyzed samples displayed 96-98% digestion efficiency.

GST-Pulldown

5 μg of GST-TIP5$_{235-741}$ were incubated with 15μl of GST beads (Glutathione Sepharose 4B, GE Healthcare) in AM100 buffer (100mM KCl, 20mM Tris-HCl pH 8.0, 5mM $MgCl_2$, 0.2mM EDTA, 1X Protease Inhibitor (Roche)) for 12-16 hours at 4°C. After two washes with EMSA buffer containing 3% Glycerol, bound GST-TIP5$_{235-741}$ was incubated with 25 nmoles of the indicated RNAs for 1h at 4°C. After a double wash with AM200 buffer (200mM KCl, 20mM Tris-HCl pH 8.0, 5mM $MgCl_2$, 0.2mM EDTA, 1X cOmplete Protease Inhibitor Cocktail (Roche)), bound GST-TIP5/RNA complexes were incubated with 0.5 μg of His-TTF1$_{1-210}$ for 2h at 4°C. Samples were then washed

three times with EBC buffer (250mM NaCl, 50mM Tris-HCl pH8.0, 0.5% NONIDET P-40, 5mM DTT, 1X cOmplete Protease Inhibitor Cocktail), run on a 12% SDS polyacrylamide gel and analyzed by Western blot with anti-GST and anti-RGS.HIS antibodies.

AP staining

Cells were fixed in 4% paraformaldehyde for 10min, washed with AP Buffer (100mM TrisCl pH 9.5, 100mM NaCl, 50mM MgCl2) and then incubated for 30 min in AP Buffer containing NBT (37 mg/ml) and 3.5µl BCIP (175 mg/ml). The staining was blocked with Tris-EDTA (Sigma) for 10min.

Whole-Transcriptome Shotgun Sequencing (RNA-Seq) and Data Analyses

Total RNA of ESCs+Control-RNA and ESCs+pRNA from two biological replicates were purified and analyzed by RNA seq. 100bp paired-end reads have been sequenced with illumina Hiseq. The reads were quality filtered and submitted to RSEM for expression quantitation (Li and Dewey, 2011). Expression counts were further analyzed with the glm method in the edgeR package to compute the significance of differential expression (Robinson et al., 2010).

Teratoma analysis

Teratoma samples were fixed in 4% buffered paraformaldehyde and embedded in paraffin. For immunostainings, 5 μm thickness paraffin sections were deparaffinized and rehydrated and subsequently subjected to the antigen retrieval (Citrate buffer pH 6.0 for 10 minutes at 110°C in rapid microwave histoprocessor, Milestone, USA). The following primary antibodies were used: anti-bIII tubulin (Sigma), anti-GFAP (DAKO) and anti-SMA (Sigma). Nuclei were stained with DAPI and slides were mounted with Fluorescent Mounting Medium (DAKO) to avoid bleaching. Images were captured with a Leica DMI 6000B Microscope and using LAS AF (Leica Application Suite Advanced Fluorescence) software. Animal experiments were performed in accordance with Swiss law and have been approved by the veterinary authorities of Zurich.

Antibodies

The following antibodies were used: anti-TIP5 (CS-090-100-Diagenode); anti-UBF (sc-13125), anti-GST (sc-459) and anti-PARP1 (sc-53643) from Santa Cruz; anti-H3K9me2 (17-648), anti-H3K9me3 (17-625), and anti-H3K27me3 (17-622) and anti-SSEA-1 from Millipore; anti-H3 (ab1791) from Abcam; anti-RGS.HIS (34610) from Qiagen. Anti-TTF1 antibody was produced with Genosphere.

List of primers

mouse rDNA CpG methylation primers		
Name	**For/Rev**	**Sequence**
rDNA promoter -165/-145	For	GACCAGTTGTTCCTTTGAGG
rDNA promoter -21/-1	Rev	ACCTATCTCCAGGTCCAATAG
rDNA coding +1/+20	For	ACTGACACGCTGTCCTTTCC
rDNA coding +111/+130	Rev	GACAGCTTCAGGCACCGCGA

mouse cDNA primers		
Name	**For/Rev**	**Sequence**
Tip5	For	AAGATGTGTGGCTACAATGG
Tip5	Rev	TCTGCACCCATCAGCTCCG
Nanog	For	AAGCAGAAGATGCGGACTGT
Nanog	Rev	ATCTGCTGGAGGCTGAGGTA
Pax6	For	GCACATGCAAACACACATGA
Pax6	Rev	ACTTGGACGGGAACTGACAC
Nestin	For	AGGCTGAGAACTCTCGCTTGC
Nestin	Rev	GGTGCTGGTCCTCTGGTATCC
Blbp	For	AGGTGGCAAAGTGGTGATCC
Blbp	Rev	TCCAACCGAACCACAGACTTAC
Minor satellites	For	CATGGAAAATGATAAAAACC
Minor satellites	Rev	CATCTAATATGTTCTACAGTGTGG
Major satellites	For	GACGACTTGAAAAATGACGAAATC
Major satellites	Rev	CATATTCCAGGTCCTTCAGTGTGC
rDNA spacer -1994/1975	For	GCAGACCGAGTTGCTGTAC
rDNA spacer -1922/1905	Rev	GGGTAGGACTTAAGCCTT
rDNA enhancer -554/-535	For	GAAGCCCTCTTGTCCCCGTC
rDNA enhancer -466/-447	Rev	GATCCAAAGCTCCAGCTGAC
rDNA promoter -165/-145	For	GACCAGTTGTTCCTTTGAGG
rDNA promoter -21/-1	Rev	ACCTATCTCCAGGTCCAATAG
45S pre-rRNA +550/570	For	CTCTTGTTCTGTGTCTGCC
45S pre-rRNA +745/765	Rev	GCCCGCTGGCAGAACGAGAAG
Line L1 ORF2	For	TTTGGGACACAATGAAAGCA
Line L1 ORF2	Rev	CTGCCGTCTACTCCTCTTGG
IAPgag	For	AGCAGGTGAAGCCACTG
IAPgag	Rev	CTTGCCACACTTAGAGC
Btg3	For	AAGGTCAGGCCTACAGATGC
Btg3	Rev	GGTCACCTTATCCAGAGCCC
Rdh10	For	GAAATCCTGCCCCCGTGTAA

Rdh10	Rev	TAGTGGTCCAGAAGTGTGCG
Oct-4	For	GGCGTTCGCTTTGGAAAGGTGTTC
Oct-4	Rev	CTCGAACCACATCCTTCTCT
Rex1	For	AGAAAGCAGGATCGCCTCAC
Rex1	Rev	AGGGAACTCGCTTCCAGAAC
Rps12	For	GAAGCTGCCAAAGCCTTAGA
Rps12	Rev	AACTGCAACCAACCACCTTC
Gapdh	For	TGCACCACCAACTGCTTAGC
Gapdh	Rev	GGCATGGACTGTGGTCATGAG

ChIP primers		
Name	**For/Rev**	**Sequence**
rDNA promoter -165/-145	For	GACCAGTTGTTCCTTTGAGG
rDNA promoter -21/-1	Rev	ACCTATCTCCAGGTCCAATAG
rDNA coding +2251/70	For	GCATCGGTGTGTCGGCATCG
rDNA coding +2346/65	Rev	CTGAGCAGTCCCACCACACC
rDNA coding +8124/145	For	GCGACCTCAGATCAGACGTGG
rDNA coding +8203/224	Rev	CTGTTCACTCGCCGTTACTGAG
Minor satellites	For	CATGGAAAATGATAAAAACC
Minor satellites	Rev	CATCTAATATGTTCTACAGTGTGG
Major satellites	For	GACGACTTGAAAAATGACGAAATC
Major satellites	Rev	CATATTCCAGGTCCTTCAGTGTGC
Evx1 TSS	For	TACACAGCATCTGGGGAGTG
Evx1 TSS	Rev	GTGTGCTGGGTTAAGGGAGA
Zfpm2 TSS	For	GGATGAAGTTCTCAGAGCTGGT
Zfpm2 TSS	Rev	GCGCGAACTTTTACACCTACTT
Dlx1 TSS	For	ATGTCTCCTTCTCCCATGTCC
Dlx1 TSS	Rev	ACTGCACGGAACTGATGTAGG
Gapdh promoter	For	GGTTGCTGTGTCACTACCGAAGAA
Gapdh promoter	Rev	AAATGGAGAAGTGTGGGTCTCCCT
Line L1 ORF2	For	TTTGGGACACAATGAAAGCA
Line L1 ORF2	Rev	CTGCCGTCTACTCCTCTTGG
IAPgag	For	AGCAGGTGAAGCCACTG
IAPgag	Rev	CTTGCCACACTTAGAGC

Supplemental References

Cammas, L., Romand, R., Fraulob, V., Mura, C., and Dolle, P. (2007). Expression of the murine retinol dehydrogenase 10 (Rdh10) gene correlates with many sites of retinoid signalling during embryogenesis and organ differentiation. Developmental dynamics : an official publication of the American Association of Anatomists *236*, 2899-2908.
Gundersen, H.J.G., Bendtsen, T.F., Korbo, L., Marcussen, N., Møller, A., Nielsen, K., Nyengaard, J.R., Pakkenberg, B., Sørensen, F.B., Vesterby, A., *et al.* (1988). Some new, simple and efficient stereological methods and their use in pathological research and diagnosis. Apmis *96*, 379-394.
Li, B., and Dewey, C.N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics *12*, 323.
Martens, J.H., O'Sullivan, R.J., Braunschweig, U., Opravil, S., Radolf, M., Steinlein, P., and Jenuwein, T. (2005). The profile of repeat-associated histone lysine methylation states in the mouse epigenome. EMBO J *24*, 800-812.
Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139-140.
Santoro, R. (2014). Analysis of chromatin composition of repetitive sequences: the ChIP-Chop assay. Methods Mol Biol *1094*, 319-328.
Santoro, R., Li, J., and Grummt, I. (2002). The nucleolar remodeling complex NoRC mediates heterochromatin formation and silencing of ribosomal gene transcription. Nat Genet *32*, 393-396.
Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., and Yamanaka, S. (2007). Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell *131*, 861-872.
Weber, F.A., Bartolomei, G., Hottiger, M.O., and Cinelli, P. (2013). Artd1/Parp1 regulates reprogramming by transcriptional regulation of Fgf4 via Sox2 ADP-ribosylation. Stem Cells *31*, 2364-2373.
West, M.J. (2012). Estimating volume in biological structures. Cold Spring Harb Protoc *2012*, 1129-1139.
Yoshida, Y., Matsuda, S., Ikematsu, N., Kawamura-Tsuzuku, J., Inazawa, J., Umemori, H., and Yamamoto, T. (1998). ANA, a novel member of Tob/BTG1 family, is expressed in the ventricular zone of the developing central nervous system. Oncogene *16*, 2687-2693.

FEBS Letters

FEBS PRESS science publishing by scientists

REVIEW ARTICLE

Challenges in the analysis of long noncoding RNA functionality

Sergio Leone[1,2] and Raffaella Santoro[1]

1 Department of Molecular Mechanisms of Disease, University of Zurich, Switzerland
2 Molecular Life Science Program, Life Science Zurich Graduate School, University of Zurich, Switzerland

Correspondence
R. Santoro, Department of Molecular Mechanisms of Disease, University of Zurich, 8057 Zurich, Switzerland
Fax: +41 44 63 55468
Tel: +41 44 63 55475
E-mail: raffaella.santoro@dmmd.uzh.ch

(Received 15 June 2016, revised 8 July 2016, accepted 11 July 2016)

doi:10.1002/1873-3468.12308

Edited by Wilhelm Just

Long noncoding RNA (lncRNA) are emerging as important regulators of diverse biological functions. Although mechanistic models are starting to emerge, it is also clear that the lncRNA field needs appropriate model systems in order to better elucidate the functions of lncRNA and their roles in both physiological and pathological conditions. The field of lncRNA is new, and the biochemical and genetic methods used to address function and mechanisms of lncRNA have only recently been developed or adapted from techniques used to investigate protein-coding genes. In this review, we discuss the strengths and weaknesses of available techniques for the analysis of chromatin-associated lncRNA and emerging models for the recruitment to specific genomic sites such as triple-helix, RNA–protein–DNA recognition and proximity-guided search models.

Keywords: chromatin; epigenetics; genome organization; lncRNA; triple helix

The G-value paradox refers to the observation that the number of protein-coding genes in different organisms does not correlate with their relative biological complexity [1]. The degree of organismal complexity among species has been proposed to better correlate with the proportion of each genome that is transcribed into noncoding RNA(ncRNA) than with the number of protein-coding genes, even when protein diversification by both alternative splicing and post-translational modifications are taken into account [2,3]. Recently, long noncoding RNA (lncRNA), a class of ncRNA longer than 200 nucleotides, have emerged as important regulators of diverse biological functions. This distinction, although based on technical aspects of RNA isolation methods, is a good compromise to distinguish this class of ncRNA from microRNA, piwi-interacting RNA, short interfering RNA, small nucleolar RNA, etc.

Long noncoding RNA have been shown to be differentially expressed across various stages of differentiation and development, which may suggest important regulatory roles in cell physiology and pathology [3,4]. Assigning functional categories to lncRNA is in itself not an easy task. Indeed, a single 1 kb lncRNA can fulfill a large number of functions, which perhaps can be active only in a defined cell type and at different

Abbreviations
CLIP, UV crosslinking and immunoprecipitation; CHART, capture hybridization analysis of RNA targets; ChIRP, chromatin isolation by RNA purification; HOTAIR, HOX antisense intergenic RNA; iCLIP, individual-nucleotide resolution CLIP; IGS-rRNA, intergenic spacer region of rRNA genes; lncRNA, long noncoding RNA; LNA, locked nucleic acids; MRE ,MSL recognition element; MSL, male-specific-lethal dosage compensation complex; ncRNA, noncoding RNA; NoRC, nucleolar remodeling complex; RNAi, RNA interference; PRC2, polycomb repressive complex 2; PAR-CLIP, photoactivatable ribonucleoside-enhanced cross-linking and immunoprecipitation; pRNA, promoter rRNA; RAP-MS, RNA antisense purification with mass spectrometry; roX1, roX2: RNA on the X 1,2; TAD, topologically associating domain; TIP5, TTF1 interacting protein 5, Baz2a; TTF1, transcription terminator factor 1; Xi, inactive X chromosome.

development stages [5]. Accordingly, despite the large amount of data published in the last few years, little is known about the function of lncRNA.

In this review, we focus exclusively on lncRNA enriched in the nucleus and, specifically, within the chromatin-associated fraction. This class of lncRNA has generally been implicated in the regulation of gene expression and establishment of epigenetic and chromatin states [6]. Although functional and mechanistic models are starting to emerge, at the core of lncRNA studies is the need for appropriate experimental systems, which should allow a better understanding of the functions of lncRNA, and their roles in both physiological and pathological conditions. In this review, we discuss the strengths and weaknesses of available techniques for the analysis of chromatin-associated lncRNA. Furthermore, we discuss the evidence that lncRNA can interact with chromatin regulatory proteins and emerging models by which lncRNA can recruit regulatory factors to specific genomic sites such as triple-helix, RNA–protein–DNA recognition and proximity-guided search models.

Toolbox to analyze lncRNA function and mechanisms in chromatin

The intimate connection between RNA and chromatin was recognized over 40 years ago in a seeding article proposing a regulatory role in gene expression of RNA found exclusively present in the nucleus and absent in the cytoplasmic fraction [7]. In the past years, a plethora of lncRNA has been identified through the application of high-throughput transcriptome analysis, and this has led to an intensive search for possible biological functions that these transcript can eventually carry. Nevertheless, molecular effects and functional significance have proven difficult to determine [8].

The field of lncRNA is new, and the biochemical and genetic methods used to address function and mechanisms of lncRNA have only recently been developed or adapted from biochemical and genetic techniques applied to study protein-coding genes. It is now clear that, to dissect the modes of action of lncRNA in the regulation of chromatin and gene expression, methods employed for the analysis of protein-coding genes must be used with caution and with appropriate controls and additional experimental strategies and methods have to be developed.

One of the most powerful techniques to study the function of a gene *in vivo* is to disrupt its expression, which can be done either using knockdown methods or through alterations of the gene locus. Downregulation of lncRNA expression through knockdown methods is usually taken as first choice to start to determine whether a given lncRNA has a function. The advantages of this method are the high efficiency of gene knockdown and the ability to easily target the gene of interest. Moreover, this method allows evaluating the role of the transcript without altering the gene locus. However, a considerable attention has to be given to the correct choice of methods to knockdown. The assessment of lncRNA localization has to be taken into consideration since depletion of lncRNA by shRNA or siRNA may be heavily impacted by their localization in the cell (Fig. 1). While this method well adapts to downregulate expression of proteins, there is a certain concern in the use of cytoplasmic RNAi machinery to knockdown nuclear lncRNA [8,9]. Instead, locked nucleic acids (LNA) or chemically modified RNA aptamers may represent a suitable method to analyze the function of nuclear lncRNA since they act by forming a DNA/RNA hybrid with the nascent RNA transcript and trigger RNase H-dependent degradation of the RNA in the nucleus [10]. Nevertheless, knockdown of low-expressed lncRNA as well as repetitive sequences might result particularly inefficient. An alternative method to destroy expression of lncRNA is through modifications of its gene locus. In this regard, and as recently discussed [8], it has to be considered that phenotypes observed from deletion of an entire genomic locus cannot be unequivocally attributed either to the loss of the lncRNA *per se* or to overlapping genomic elements. Moreover, it cannot be excluded that the act of transcription *per se* through a lncRNA locus can also generate changes in chromatin structure. Therefore, additional experimental evidences that can discriminate the role of the genomic locus from that of its RNA products are needed to confirm or reject the hypothesis of functionality of a lncRNA as molecular species. In this context, the recent advances in genome editing using designer site-specific nucleases such as CRISPR/Cas9 and TALENs could be very helpful to allow an adequate dissection of lncRNA function and mechanisms *in vivo* (Fig. 1). These tools can allow for instance the fine regulation of lncRNA transcription by targeting directly regulatory regions without affecting the transcriptional regulation of the neighboring genes and help to perform direct perturbation experiments such as loss-of-function and gain-of-function.

The analysis of lncRNA function often relies on the identification of interacting proteins or nucleic acids by RNA–protein immunoprecipitation. However, many assays generally used for protein analyses might be prone to nonspecific binding in the context of

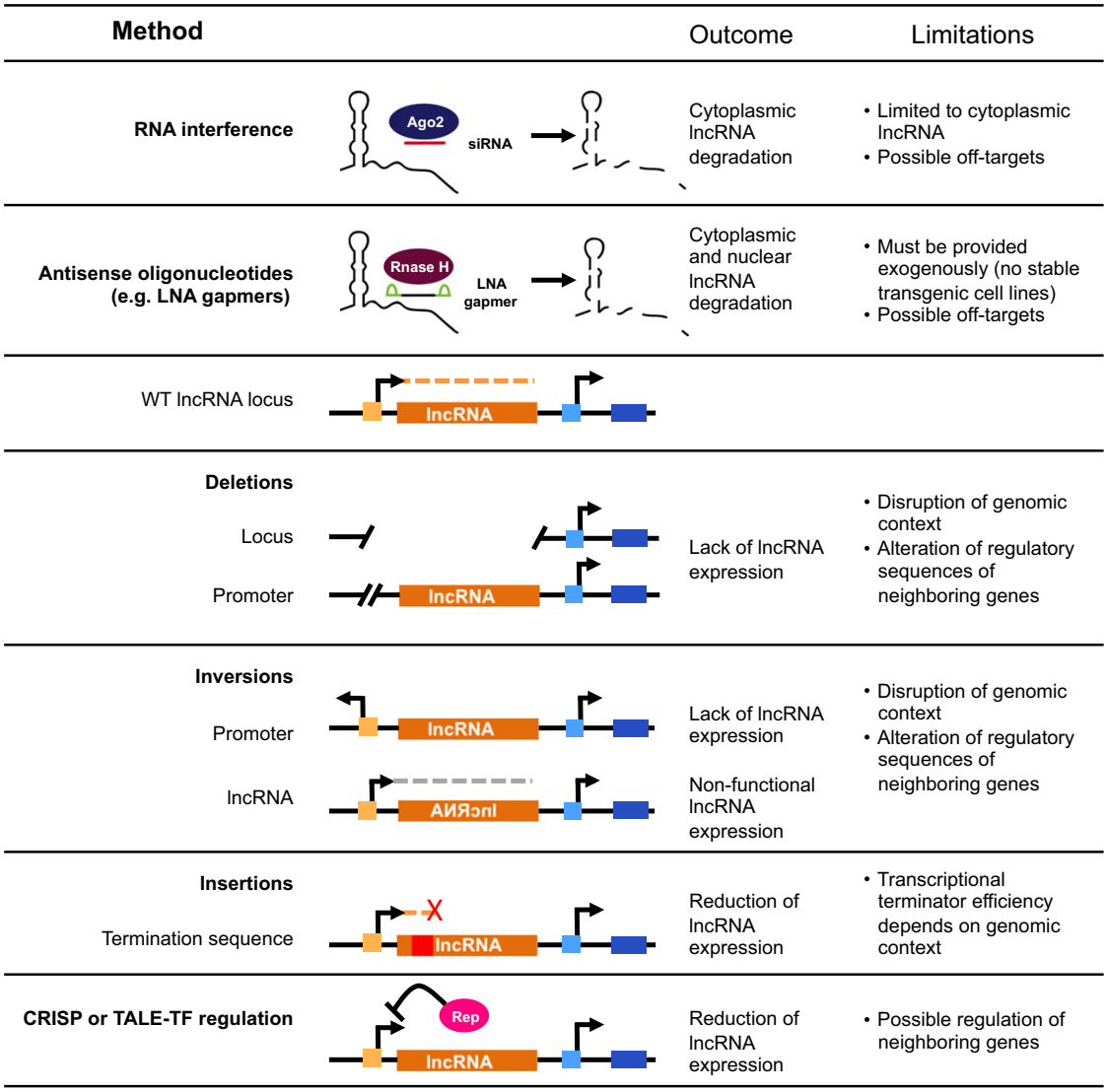
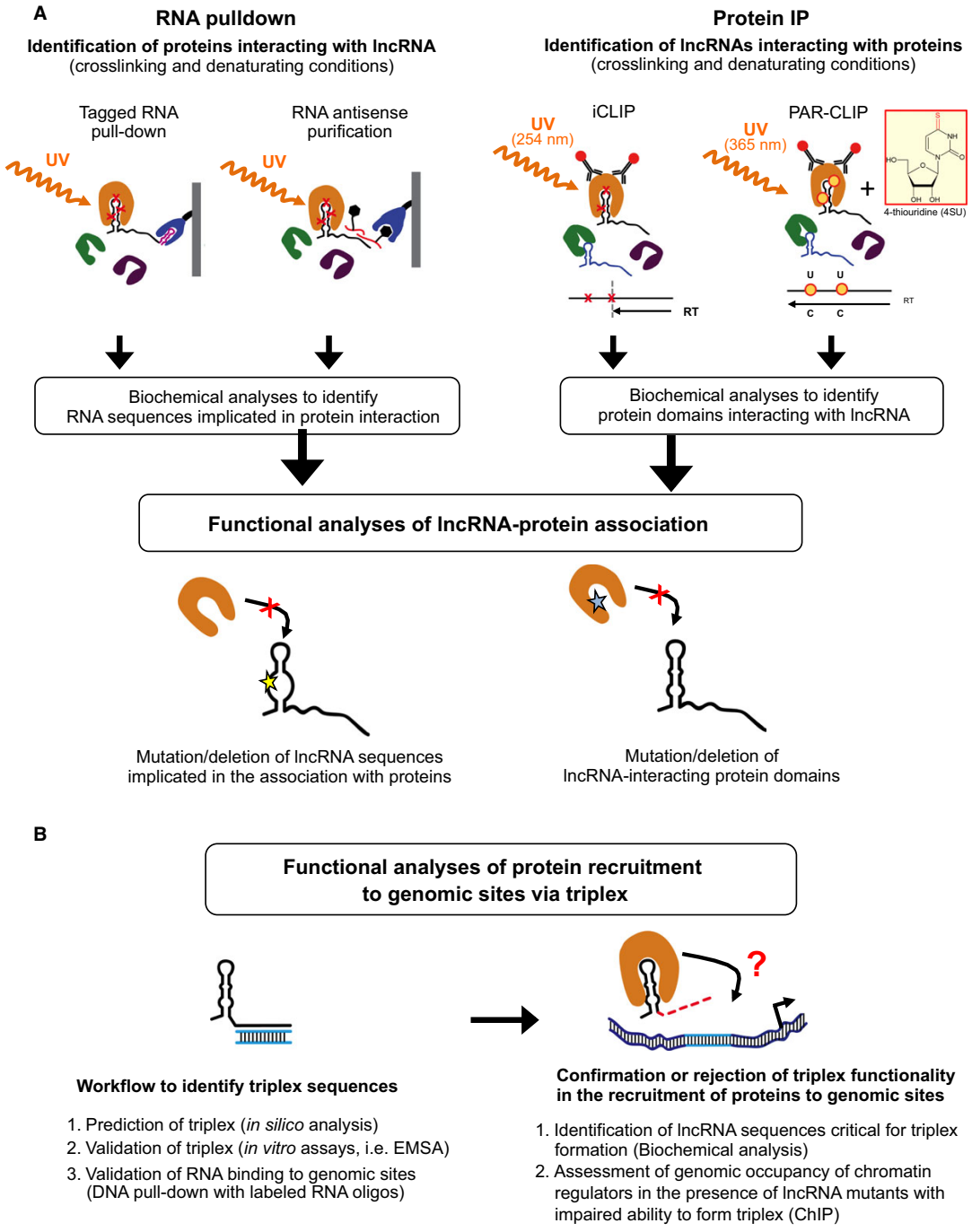


Fig. 1. Strategies to analyze lncRNA loss of function. Strategies to alter lncRNA expression. The lncRNA locus is indicated in orange, neighboring protein-coding gene in blue, and the process of transcription by dotted lines. Knockdown of lncRNA can be achieved by RNA interference (RNAi), a process activated by dsRNA species delivered to the cytoplasm of cells. Downregulation of nuclear lncRNA can be achieved using antisense oligonucleotides (LNA or chemically modified RNA aptamers), which bind to nascent transcripts forming a DNA/RNA hybrid, triggering RNase H-dependent degradation of the RNA in the nucleus. Examples how to alter lncRNA gene locus are also depicted. Another possibility is genome engineering using dCas9 and TALENs fused to a repressor (Rep), which can establish transcriptional silencing of lncRNA genes. Outcome and limitations of the method chosen to alter lncRNA expression are described.

RNA–protein interaction. Although methods using 'native purification' (purification of RNA–protein complexes under physiological conditions) have been often used in the past due to the advantage to preserve the native complexes present in the cell, they have several limitations, including the potential to identify RNA–protein interactions that form only during the preparation of cell extracts but do not occur in the cell [11]. The use of new techniques designed to identify direct lncRNA-interactors by cross-linking and subsequent purification upon denaturing conditions to remove nonspecific interactions (iCLIP, PAR-CLIP) is



clearly necessary to demonstrate the interaction of a protein with a lncRNA in cells [12–14] (Fig. 2A). Along this line, analyses based on purification by antisense biotinylated probes have been recently developed to study RNA–protein interactions also in the context of chromatin (RAP-MS, ChIRP, CHART)

Fig. 2. Strategies to analyze the function of lncRNA association with proteins and DNA. (A) Methods to identify lncRNA–protein interactions are described. RNA-based approaches (RNA pull-downs) allow to identify proteins interacting with a specific RNA. *In vivo* cross-linking (red Xs) and purification under denaturing conditions ensure the isolation of proteins interacting with lncRNA in the cell. lncRNA–protein complexes can be efficiently isolated using ectopically expressed RNA tagged with a specific RNA sequence (purple hairpin) and subsequent purification through a RNA affinity matrix. Another efficient method involves the hybridization of long antisense biotinylated probes and purification by streptavidin resins. Protein-based approaches (protein IP) rely on antibodies able to immunoprecipitate (IP) the RNA-interacting protein of interest. Also in this case cross-linking strategies are recommended to catch-specific interactions. Techniques like iCLIP or PAR-CLIP allow to identify not only RNA-interacting proteins but also to map RNA sequences in contact with proteins. To determine the functionality of lncRNA-protein association, we suggest to perform biochemical analyses that allow identifying RNA sequences or protein domains involved in lncRNA–protein interactions. The results obtained by this analysis will instruct in the engineering of mutations or deletion at lncRNAs or proteins that serve to analyze the role of lncRNA–protein association in the cell. (B) Workflow to determine the function of triplex in the recruitment of protein complexes to specific genomic sites. Upon the identification of RNA sequences with ability to form triplex, biochemical analyses should be performed to determine which lncRNA sequences are critical for triplex formation. The engineering of lncRNA mutants with impaired ability to form triplex will determine whether the association of chromatin regulators with defined genomic sites depends on lncRNA-triplex. iCLIP individual-nucleotide resolution cross-linking and immunoprecipitation; PAR-CLIP photoactivatable-ribonucleoside-enhanced cross-linking and immunoprecipitation.

[15–17]. Finally, the identification of a lncRNA-interactor should be accompanied by biochemical assays to measure the affinity of a defined protein for a lncRNA and determine whether the association depends on the RNA sequence or its structure. Indeed, the lncRNA field will benefit on the analysis of lncRNA mutants with impaired ability to associate with a defined chromatin regulator or target genes to verify the functionality of lncRNA–protein interaction and/or their association with genomic sites (Fig. 2).

Thus, it is clear that to fully resolve the true *in vivo* functions of lncRNA, it is necessary to take into account the strengths and weaknesses of the available techniques.

lncRNA as scaffold for protein complexes

A major recurrent theme in lncRNA biology is the ability to act as scaffold for the assembly of chromatin regulator complexes, which in turn might serve a guider function for the recruitment at specific genomic loci. A RNA scaffold may have a selectable advantage over protein for many applications. A typical RNA 'arm' of 50 base pairs (with bulges and internal loops) extends for 13 nm, whereas a 50-amino-acid alpha helix extends for 7.5 nm [5]. Thus, lncRNA may easily bind multiple proteins and act as scaffold for the assembly of chromatin regulatory complexes and other protein complexes.

Initial evidence for the role of lncRNA in gene regulation came from studies of mammalian X-chromosome inactivation, a process regulated by the *Xist* lncRNA, which is transcribed exclusively from the inactive X-chromosome (Xi) [18,19]. Genetic deletion of *Xist* prevents X chromosome inactivation [20], and transgenic *Xist* RNA caused long-range transcriptional repression in cis [21]. The role of *Xist* in silencing depends on a conserved repeat sequence, the A-repeat domain, whose deletion prevents transcriptional silencing without affecting its association with chromatin and spreading over the X-chromosome [22]. A direct association of polycomb repressive complex 2 (PRC2) with lncRNA has been proposed as a mechanism for recruitment and establishment of H3 lysine 27 methylation (H3K27me3) at Xi [23]. Indeed, the timing of PRC2 recruitment on X chromosome tightly coincides with the induction of *Xist* during development and is strictly dependent on continuous *Xist* RNA transcription [24–26]. Initial analyses obtained by RIP and pull-down assays with biotinylated RNA have supported a model where EZH2, a component of PRC2, directly interacts with *Xist* [23,27]. However, recent results have started to challenge the current model based on *Xist*-PRC2 association due to the promiscuous binding of PRC2 to RNA substrates, inhibition of EHZ2 activity by RNA and spatial separation of *Xist* RNA and polycomb proteins as revealed by super-resolution microscopy [28–30]. At the basis of this debate is clearly the weakness of 'native purification' methods, which have been initially used to detect the *Xist*-PRC2 interaction. Recently, a novel method (RAP-MS) based on long biotinylated antisense probes, which form very stable RNA–DNA hybrids, has been used to purify lncRNA complexes in denaturing and reducing conditions [15]. Ten proteins were identified that specifically associate with *Xist*, three of these proteins (SHARP, SAF-A and LBR) are required for *Xist*-mediated transcriptional silencing. SHARP, which interacts with the SMRT corepressor that activates HDAC3, is required for *Xist*-mediated recruitment of PRC2 across the X chromosome.

Remarkably, this and another study using similar stringent conditions to identify proteins directly interacting with *Xist* failed to detect PRC2 [15,31].

HOX antisense intergenic RNA (HOTAIR), a lncRNA residing in the HOXC locus, is another important example of the intimate link between lncRNA and gene silencing [32]. Depletion of HOTAIR by siRNA led to transcriptional activation of the *HOXD* locus on chromosome 2 spanning over 40 kb, suggesting that HOTAIR is required to maintain a transcriptionally silent chromosomal domain in *trans*. Accordingly, enforced expression of HOTAIR in epithelial cancer cells induced genome-wide retargeting of PRC2, leading to altered H3K27me3 and gene expression [33]. HOTAIR has been proposed to serve as modular scaffold of histone modification complexes, including PRC2, and thereby specifies the pattern of histone modifications on target genes [33].

Another example representing the regulation of chromatin state by lncRNA is pRNA, a 250–300 nt lncRNA that is the product of processing of a 2-kb-long transcript derived from the intergenic spacer region of rRNA genes (IGS-rRNA) [34,35]. The formation of heterochromatin at rRNA genes, which are located in the nucleolus, depends on pRNA: LNA-mediated depletion of pRNA prevents rRNA gene silencing [36], whereas ectopic expression of mature pRNA induces heterochromation formation and silencing [34]. pRNA associates with TIP5 (Baz2a), a component of the nucleolar remodeling complex NoRC, that previous studies have implicated in epigenetic silencing of rRNA genes [37]. A stem loop structure within pRNA is required for the association with TIP5: point mutations destroying pRNA stem loop impair the association of TIP5 with pRNA and the recruitment of TIP5 to rRNA genes, whereas a compensatory mutation that restores loop structure is sufficient to guide TIP5 to rRNA genes [34,38]. pRNA serves as modular scaffold to promote the association of TIP5 with other factors such as poly (ADP-Ribose) Polymerase 1 (PARP1) and transcription terminator factor 1 (TTF1) [34,39]. Importantly, although the unprocessed transcript IGS-rRNA abolishes the association of TIP5 with TTF1, mature pRNA promotes this interaction that serves to dock the complex at the promoter of rRNA genes and to establish heterochromatin during embryonic stem cell differentiation [34]. These results suggest that the same lncRNA can prevent or promote protein complex assembly and its processing controls the switch between these functions. Thus, lncRNA processing may represent an additional level of lncRNA regulation by modulating distinct features of the same lncRNA such as the assembly of different protein complexes.

Two ncRNA, *roX1* (3.7 kb) and *roX2* (0.5 kb), are essential for gene dosage compensation in *Drosophila* [40]. These RNA associate with the ribonucleoprotein male-specific lethal dosage compensation complex (MSL), which comprises at least five proteins, MSL1, MSL2, MSL3, MOF (males-absent-on-the first), and MLE (maleless). The MSL complex assembles exclusively in male flies, binds to hundreds of sites on the male X chromosome, and increases transcription from X-linked genes. Incorporation of roX1/2 into the MSL complex is mediated by the RNA helicase MLE and involves transient RNA-mediated interactions with the core MSL complex. Biochemical and genetic evidences have shown that *roX1* and *roX2* provide a scaffold for arrangement of the MSL proteins and is prerequisite for MLE localization to the male X chromosome [41,42].

lncRNA recruitment to genomic DNA

Several mechanisms have been proposed for lncRNA-mediated recruitment of chromatin regulators at defined regions of the genome. lncRNA (a) can interact with DNA through direct nucleic acid hybridization, (b) can physically interact with DNA-binding proteins, or (c) can exploit the 3D conformation of the nucleus to search for targets. These mechanisms do not exclude each other and may act together to target lncRNA to specific genomic sites.

The triple-helix formation involves a double-stranded nucleic acid such as duplex DNA and a single-stranded nucleic acid such as RNA [43]. Triplex-helices are formed by sequence-specific binding rules: the single-stranded nucleic acid binds in the major groove of the targeted duplex through sequence-specific recognition of a polypurine·polypyrimidine sequence (Hoogsteen or reverse Hoogsteen base-pairing) [44]. Initial immunofluorescence studies using triple-helix-specific monoclonal antibodies have suggested the existence of triplex structures in the nucleus of insects, nematodes, and mammals [45]. Triple helix formation has been often evoked as a potential mechanism of lncRNA-mediated recruitment of chromatin regulators to specific genomic sites. However, evidences for this mechanism are generally circumstantial and still lack a direct and robust experimental proof. Indeed, the solely demonstration of the ability of a given RNA to form triple-helix *in vitro* cannot prove that this is the mechanism by which protein complexes are recruited to specific genomic sequences. Functionality of lncRNA triple-helix

should be carefully proven and confirmation or rejection of hypothesis should be exploited through perturbation experiments such as mutations in lncRNA sequences implicated in triplex formation to validate the binding to genomic sites and the recruitment of chromatin regulators (Fig. 2B).

Fendrr is a lncRNA proposed to act via dsDNA–RNA triplex formation at target regulatory elements and to increase PRC2 occupancy at these sites [46]. This model was proposed based on the presence of a purine-pyrimidine 40-nucleotide stretch motif that is favorable for triplex formation. The binding of a synthetic RNA oligonucleotide to double-stranded *Foxf1* and *Pitx2* promoter fragments in an *in vitro*-binding assay in the presence of RNase H, which specifically cleaves RNA–DNA hybrids, led to propose that Fendrr anchors PRC2 at its target promoters in cells through triplex formation [46]. However, additional investigations are required to confirm this triple-helix model, in particular, whether recruitment of PRC2 to target genes in cells requires *Fendrr* sequences implicated in the formation of triple-helix *in vitro*.

The lncRNA *PARTICLE* was shown to afford both a cytosolic scaffold for the tumor suppressor methionine adenosyltransferase (MAT2A), and a nuclear genetic platform for transcriptional repression [47]. *PARTICLE* represses *MAT2A* and it was proposed to recruit the transcription-repressive complex proteins G9a and SUZ12 (subunit of PRC2) through triplex formation. This model was based on an *in vitro*-binding assay using surface plasmon resonance technology that determined the ability of *PARTICLE* to form triple-helix with a synthetic dsDNA *MATA2* sequence. However, whether transcriptional silencing and recruitment of repressors to *MATA2* depends on *PARTICLE* sequences implicated in triple-helix was not further investigated.

The lncRNA *MEG3* regulates the TGF-β pathway genes and it was proposed to guide chromatin repressor complexes (including PRC2) through formation of RNA–DNA triplex structures [48]. EMSA assays and circular dichroism spectroscopy analyses revealed that several *MEG3* sequences have the ability to form triple-helix with *MEG3*-bound genomic sites enriched in GA sequences. Interestingly, *MEG3* sequences able to form triplex do not require a perfect identity with the target DNA sequences. The association of one of these RNA with *MEG3*-bound genomic sites was further determined through pulldown assays in the presence of RNase H using biotin-labeled oligos, which were transfected or incubated with nuclei isolated from BT-549 cells. However, as in the studies described above, further investigations are required to determine whether the recruitment of repressor complexes is mediated by triplex formation. An interesting part of this work was the analysis of triplex using an anti-triplex dA.2rU antibody, which was raised against the triplex derived from homopolymeric nucleic acids (poly(rU).poly(dA).poly(rU)). Immunostaining with this antibody revealed a signal that was higher in the nuclear compartment than in the cytoplasm and resistant to RNase H treatment. Dot-blot analysis showed that this antibody recognized Poly (rU):Poly (dA)-Poly (dT) and $[(dCTT)]_7$: $[(dGAA)]_7$ – $[(dCTT)]_7$, indicating that antibody reactivity is not restricted to three-stranded configurations assembled with homopolymeric nucleic acids but also can recognize triplex DNA made with poly-purine/poly-pyrimidine sequences. Interestingly, this antibody also showed a strong affinity for Poly (rU)-Poly (dA). Certainly, an in-deep analysis of the specificity of this antibody using a larger panel of RNA:dsDNA sequences would have been more informative. Moreover, there is a certain lack of information concerning the origin of this antibody since its description was accompanied by a citation of a work published 40 years ago [49] and there is no indication whether this antibody was purchased from a commercial source, or it was raised in-house. Clearly, if this antibody turns out to be highly specific for triple-helix as described by Mondal *et al.* [48], it will be an important tool for the analysis of lncRNA functionality.

Another example of triplex-mediated recruitment of chromatin regulators is pRNA. A previous work has shown that T0 sequences located at the 5′- terminus of pRNA can form triple helix at the promoter of rRNA genes [50]. While RNase H, EMSA, and psoralen-based assays demonstrate that the binding of pRNA to the promoter of rRNA genes is neither via formation of RNA–DNA hybrids nor due to mediating proteins [50], the sequence implicated in this putative triple helix greatly deviates from the common triplex model (G, A; purine motif) and, hence, further investigations into the proposed triplex formation are required [44]. The same work reported that the formation of triple-helix is required for the recruitment of the *de novo* DNA methyltransferase DNMT3b to rRNA genes [50]. Remarkably, the pRNA triple-helix model is often reported in reviews as a mechanism for the recruitment of NoRC complex to rRNA genes [51–54]. However, there is no experimental evidence supporting this mechanism. First, the original report of pRNA triple-helix did not analyze whether NoRC is recruited to rRNA genes via pRNA triplex. [50]. Second, a recent work showed that recruitment of NoRC to rRNA genes does not require pRNA

sequences implicated in triple-helix: deletion or mutation of T0 sequences in pRNA efficiently guide NoRC to rRNA genes and initiate *de novo* methylation and formation of heterochromatin, a result that is also consistent with previous studies showing the association of DNMT3b with TIP5, a subunit of NoRC [34,37]. Recent results have proposed that pRNA-mediated targeting of TIP5 occurs through DNA–protein recognition rules [34]. TTF1, a nucleolar protein that binds to terminator (T) elements, including the T0 sequences at rRNA gene promoter, was previously identified as an interactor of TIP5 (TTF1 interacting protein 5) and required for the association of TIP5 with rRNA gene promoter [55–57]. Biochemical analyses revealed that the association of TIP5 with TTF1 tightly depends on the stem-loop structure of pRNA, suggesting that the interaction of lncRNA with sequence specific DNA-binding proteins is a driver for recruitment of chromatin regulators to defined genomic sites [34]. Recruitment mediated by the RNA–protein–DNA module has also been proposed for YY1, a 'bivalent' protein, capable of binding both *Xist* RNA and DNA, an association critical to tether *Xist* RNA to the future Xi and to nucleate the coating of *Xist* RNA along the Xi [58]. A similar model can be also applied to the recruitment of *Drosophila* MSL complex, which includes *rox1* and *rox2*, to MSL recognition elements (MREs) that is dependent on the interaction with the CLAMP, a DNA-binding protein that recognizes MREs [59]. A RNA–protein–DNA module has also been proposed to explain the localization of *Xist* and *Firre*: both lncRNA are thought to interact with the hnRNPU/SAF-A DNA-binding protein, which is required for their localization to DNA [11,60,61]. In the light of these results, we think that the function of triplex for the recruitment of chromatin regulators to specific genomic sites remains still circumstantial. The field of lncRNA would benefit on the analysis of lncRNA mutants with impaired ability for triplex formation to verify the requirement of triplex structures for the recruitment of protein complexes to specific genomic loci.

The 3D conformation of the nucleus has been recently proposed as a potentially general mechanism by which lncRNA search for genomic sites [11]. Several lncRNA have recently been shown to use spatial proximity to identify target sites. Chromosomal looping brings *HOTTIP* RNA in close proximity to the 5′ *HOXA* genes and *HOTTIP* yields a broad domain of H3K4me3 and transcription activation at 5′ *HOXA* locus through the association with WDR5-MLL complexes [62]. Recently, it was shown that *Xist* coats the X chromosome by searching in three dimensions: first it localizes to genomic sites that are in close spatial proximity to its own transcription locus and then spreads to newly accessible locations [63,64]. Similarly, in *Drosophila*, high affinity sites, landing platforms of MSL complex, have been found to be enriched around topologically associating domain (TAD) boundaries on the X chromosome and harbor more long-range contacts in a sex-independent manner, which might provide an advantageous location for the MSL complex to spread to spatially close regions and induce dosage compensation [65].

It is important also to take into consideration the possibility that the proximity-guided search and lncRNA localization can occur across different chromosomes, which are in close spatial proximity in the nucleus. This mode of action might be used by *HOTAIR*, which is transcribed from the HoxC locus but regulates the expression of genes in the HoxD locus, present on a different chromosome [32]. Another example is provided by the nucleolus, which forms a specific domain where rRNA gene loci, which are spread across multiple chromosomes, coalesce into a spatially organized compartment. The spatial close proximity of rRNA genes might facilitate pRNA, which is transcribed from active rRNA genes during early S-phase, to guide in trans the NoRC complex to establish heterochromatin of silent rRNA genes after their replication in mid-late S phase [35,66]. Finally, *Firre* localizes across five distinct transchromosomal loci, which reside in spatial proximity to the *Firre* genomic locus on the X chromosome [60]. Both genetic deletion of the Firre locus and knockdown of Firre-interacting protein hnRNPU resulted in loss of colocalization of these transchromosomal interacting loci, suggesting a model in which lncRNA can interface with and modulate nuclear architecture across chromosomes [60].

lncRNA are generally less abundant than mRNA. Therefore, the 3D genome organization might facilitate the recruitment of lncRNA to target genes if these are located closed to lncRNA transcription locus where transcript concentration is higher [67]. However, proximity alone is not sufficient to explain interactions, otherwise any genomic region in close proximity of lncRNA genomic loci will be regulated. Indeed, the *Firre* lncRNA interacts with specific DNA sites that are in spatial proximity to the *Firre* locus, but does not interact with all sites in spatial proximity [60]. The combination of other mechanisms, such as DNA-binding interactions, must therefore contribute for proper localization of the lncRNA to specific genomic loci. Future studies will be required to dissect the interplay of lncRNA and the dynamic nuclear organization at the molecular level. The systematic perturbation of lncRNA, including deletion/mutations of protein and/or DNA-binding sequences, will help to determine

their roles in the establishment of chromatin architecture and epigenetic states.

Conclusions and perspectives

Multiple lines of evidences increasingly link mutations and dysregulations of lncRNA to diverse human diseases, including cancer [68]. The recent application of next-generation sequencing to a growing number of cancer transcriptomes has indeed revealed thousands of lncRNA whose aberrant expression is associated with different cancer types [4]. LncRNA expression has been correlated with distinct gene sets that influence cell cycle regulation, survival, immune response, migration, genomic stability, or pluripotency, which determine the transformed phenotype of cancer cells (we refer readers to several recent reviews that cover the classification of cancer-related lncRNA in greater details [4,69]). The tissue-specific expression of lncRNA and the correlation with disease phenotype might hold the promise of tailored therapeutic applications that fine-tune the regulatory networks of cancer cells in a highly cancer type-specific manner [4]. However, an absolute requirement is a careful characterization of individual lncRNA in terms of functions and mechanisms, both in physiological and pathological conditions. Deciphering the mechanistic diversity of the many lncRNA, their influence on nuclear architecture and their cellular roles will be important to understand disease and facilitate the establishment of diagnostic and therapeutic strategies.

In this review we have highlighted how the analysis of lncRNA to assign the exact mode of action and functional roles is more complex and difficult than that of protein-coding genes. The study of lncRNA is still in its infancy and to resolve the true *in vivo* functions of lncRNA, the strengths and weaknesses of available techniques have to be carefully weighed. The field of lncRNA biology would benefit greatly from the development of additional experimental strategies and dedicated technologies that clearly dissect the pathways mediated by lncRNA as molecular species and the molecular mechanisms of lncRNA function. We indeed strongly agree with the conclusion of Schmitz *et al*. [69] that 'It is now up to the responsibility of journal editors and reviewers to enforce the high standards of investigation that this new and exciting field of research deserves'.

Acknowledgement

This work was supported by the Swiss National Science Foundation (310003A-152854) and Krebsforschungs Schweiz (KFS-3497-08-2014).

Author contributions

SL and RS wrote the article and prepared figures.

References

1 Hahn MW and Wray GA (2002) The g-value paradox. *Evol Dev* **4**, 73–75.
2 Taft RJ, Pheasant M and Mattick JS (2007) The relationship between non-protein-coding DNA and eukaryotic complexity. *BioEssays* **29**, 288–299.
3 Fatica A and Bozzoni I (2014) Long non-coding RNAs: new players in cell differentiation and development. *Nat Rev Genet* **15**, 7–21.
4 Huarte M (2015) The emerging role of lncRNAs in cancer. *Nat Med* **21**, 1253–1261.
5 Cech TR and Steitz JA (2014) The noncoding RNA revolution-trashing old rules to forge new ones. *Cell* **157**, 77–94.
6 Guttman M and Rinn JL (2012) Modular regulatory principles of large non-coding RNAs. *Nature* **482**, 339–346.
7 Britten RJ and Davidson EH (1969) Gene regulation for higher cells: a theory. *Science* **165**, 349–357.
8 Bassett AR, Akhtar A, Barlow DP, Bird AP, Brockdorff N, Duboule D, Ephrussi A, Ferguson-Smith AC, Gingeras TR, Haerty W *et al*. (2014) Considerations when investigating lncRNA function in vivo. *Elife* **3**, e03058.
9 Zhang Q, Chen CY, Yedavalli VS and Jeang KT (2013) NEAT1 long noncoding RNA and paraspeckle bodies modulate HIV-1 posttranscriptional expression. *MBio* **4**, e00596-12.
10 Bennett CF and Swayze EE (2010) RNA targeting therapeutics: molecular mechanisms of antisense oligonucleotides as a therapeutic platform. *Annu Rev Pharmacol Toxicol* **50**, 259–293.
11 Quinodoz S and Guttman M (2014) Long noncoding RNAs: an emerging link between gene regulation and nuclear organization. *Trends Cell Biol* **24**, 651–663.
12 Sugimoto Y, Konig J, Hussain S, Zupan B, Curk T, Frye M and Ule J (2012) Analysis of CLIP and iCLIP methods for nucleotide-resolution studies of protein-RNA interactions. *Genome Biol* **13**, R67.
13 Ascano M, Hafner M, Cekan P, Gerstberger S and Tuschl T (2012) Identification of RNA-protein interaction networks using PAR-CLIP. *Wiley Interdiscip Rev RNA* **3**, 159–177.
14 Konig J, Zarnack K, Luscombe NM and Ule J (2011) Protein-RNA interactions: new genomic technologies and perspectives. *Nat Rev Genet* **13**, 77–83.
15 McHugh CA, Chen CK, Chow A, Surka CF, Tran C, McDonel P, Pandya-Jones A, Blanco M, Burghard C, Moradian A *et al*. (2015) The Xist lncRNA interacts

directly with SHARP to silence transcription through HDAC3. *Nature* **521**, 232–236.
16 Chu C, Qu K, Zhong FL, Artandi SE and Chang HY (2011) Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions. *Mol Cell* **44**, 667–678.
17 Simon MD, Wang CI, Kharchenko PV, West JA, Chapman BA, Alekseyenko AA, Borowsky ML, Kuroda MI and Kingston RE (2011) The genomic binding sites of a noncoding RNA. *Proc Natl Acad Sci USA* **108**, 20497–20502.
18 Brown CJ, Ballabio A, Rupert JL, Lafreniere RG, Grompe M, Tonlorenzi R and Willard HF (1991) A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. *Nature* **349**, 38–44.
19 Brockdorff N, Ashworth A, Kay GF, Cooper P, Smith S, McCabe VM, Norris DP, Penny GD, Patel D and Rastan S (1991) Conservation of position and exclusive expression of mouse Xist from the inactive X chromosome. *Nature* **351**, 329–331.
20 Penny GD, Kay GF, Sheardown SA, Rastan S and Brockdorff N (1996) Requirement for Xist in X chromosome inactivation. *Nature* **379**, 131–137.
21 Wutz A and Jaenisch R (2000) A shift from reversible to irreversible X inactivation is triggered during ES cell differentiation. *Mol Cell* **5**, 695–705.
22 Wutz A, Rasmussen TP and Jaenisch R (2002) Chromosomal silencing and localization are mediated by different domains of Xist RNA. *Nat Genet* **30**, 167–174.
23 Zhao J, Sun BK, Erwin JA, Song JJ and Lee JT (2008) Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science* **322**, 750–756.
24 Silva J, Mak W, Zvetkova I, Appanah R, Nesterova TB, Webster Z, Peters AH, Jenuwein T, Otte AP and Brockdorff N (2003) Establishment of histone h3 methylation on the inactive X chromosome requires transient recruitment of Eed-Enx1 polycomb group complexes. *Dev Cell* **4**, 481–495.
25 Plath K, Fang J, Mlynarczyk-Evans SK, Cao R, Worringer KA, Wang H, de la Cruz CC, Otte AP, Panning B and Zhang Y (2003) Role of histone H3 lysine 27 methylation in X inactivation. *Science* **300**, 131–135.
26 Kohlmaier A, Savarese F, Lachner M, Martens J, Jenuwein T and Wutz A (2004) A chromosomal memory triggered by Xist regulates histone methylation in X inactivation. *PLoS Biol* **2**, E171.
27 Zhao J, Ohsumi TK, Kung JT, Ogawa Y, Grau DJ, Sarma K, Song JJ, Kingston RE, Borowsky M and Lee JT (2010) Genome-wide identification of polycomb-associated RNAs by RIP-seq. *Mol Cell* **40**, 939–953.
28 Cerase A, Smeets D, Tang YA, Gdula M, Kraus F, Spivakov M, Moindrot B, Leleu M, Tattermusch A, Demmerle J *et al.* (2014) Spatial separation of Xist RNA and polycomb proteins revealed by superresolution microscopy. *Proc Natl Acad Sci USA* **111**, 2235–2240.
29 Cifuentes-Rojas C, Hernandez AJ, Sarma K and Lee JT (2014) Regulatory interactions between RNA and polycomb repressive complex 2. *Mol Cell* **55**, 171–185.
30 Davidovich C, Zheng L, Goodrich KJ and Cech TR (2013) Promiscuous RNA binding by polycomb repressive complex 2. *Nat Struct Mol Biol* **20**, 1250–1257.
31 Chu C, Zhang QC, da Rocha ST, Flynn RA, Bharadwaj M, Calabrese JM, Magnuson T, Heard E and Chang HY (2015) Systematic discovery of Xist RNA binding proteins. *Cell* **161**, 404–416.
32 Rinn JL, Kertesz M, Wang JK, Squazzo SL, Xu X, Brugmann SA, Goodnough LH, Helms JA, Farnham PJ, Segal E *et al.* (2007) Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell* **129**, 1311–1323.
33 Gupta RA, Shah N, Wang KC, Kim J, Horlings HM, Wong DJ, Tsai MC, Hung T, Argani P, Rinn JL *et al.* (2010) Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature* **464**, 1071–1076.
34 Savic N, Bar D, Leone S, Frommel SC, Weber FA, Vollenweider E, Ferrari E, Ziegler U, Kaech A, Shakhova O *et al.* (2014) lncRNA maturation to initiate heterochromatin formation in the nucleolus is required for exit from pluripotency in ESCs. *Cell Stem Cell* **15**, 720–734.
35 Santoro R, Schmitz KM, Sandoval J and Grummt I (2010) Intergenic transcripts originating from a subclass of ribosomal DNA repeats silence ribosomal RNA genes in trans. *EMBO Rep* **11**, 52–58.
36 Mayer C, Schmitz KM, Li J, Grummt I and Santoro R (2006) Intergenic transcripts regulate the epigenetic state of rRNA genes. *Mol Cell* **22**, 351–361.
37 Santoro R, Li J and Grummt I (2002) The nucleolar remodeling complex NoRC mediates heterochromatin formation and silencing of ribosomal gene transcription. *Nat Genet* **32**, 393–396.
38 Mayer C, Neubert M and Grummt I (2008) The structure of NoRC-associated RNA is crucial for targeting the chromatin remodelling complex NoRC to the nucleolus. *EMBO Rep* **9**, 774–780.
39 Guetg C, Scheifele F, Rosenthal F, Hottiger MO and Santoro R (2012) Inheritance of silent rDNA chromatin is mediated by PARP1 via noncoding RNA. *Mol Cell* **45**, 790–800.

40 Keller CI and Akhtar A (2015) The MSL complex: juggling RNA-protein interactions for dosage compensation and beyond. *Curr Opin Genet Dev* **31**, 1–11.
41 Prabu JR, Muller M, Thomae AW, Schussler S, Bonneau F, Becker PB and Conti E (2015) Structure of the RNA helicase MLE reveals the molecular mechanisms for uridine specificity and RNA-ATP coupling. *Mol Cell* **60**, 487–499.
42 Ilik IA, Quinn JJ, Georgiev P, Tavares-Cadete F, Maticzka D, Toscano S, Wan Y, Spitale RC, Luscombe N, Backofen R *et al*. (2013) Tandem stem-loops in roX RNAs act together to mediate X chromosome dosage compensation in *Drosophila*. *Mol Cell* **51**, 156–173.
43 Buske FA, Bauer DC, Mattick JS and Bailey TL (2012) Triplexator: detecting nucleic acid triple helices in genomic and transcriptomic data. *Genome Res* **22**, 1372–1381.
44 Buske FA, Mattick JS and Bailey TL (2011) Potential in vivo roles of nucleic acid triple-helices. *RNA Biol* **8**, 427–439.
45 Lee JS, Burkholder GD, Latimer LJ, Haug BL and Braun RP (1987) A monoclonal antibody to triplex DNA binds to eucaryotic chromosomes. *Nucleic Acids Res* **15**, 1047–1061.
46 Grote P, Wittler L, Hendrix D, Koch F, Wahrisch S, Beisaw A, Macura K, Blass G, Kellis M, Werber M *et al*. (2013) The tissue-specific lncRNA Fendrr is an essential regulator of heart and body wall development in the mouse. *Dev Cell* **24**, 206–214.
47 O'Leary VB, Ovsepian SV, Carrascosa LG, Buske FA, Radulovic V, Niyazi M, Moertl S, Trau M, Atkinson MJ and Anastasov N (2015) PARTICLE, a triplex-forming long ncRNA, regulates locus-specific methylation in response to low-dose irradiation. *Cell Rep* **11**, 474–485.
48 Mondal T, Subhash S, Vaid R, Enroth S, Uday S, Reinius B, Mitra S, Mohammed A, James AR, Hoberg E *et al*. (2015) MEG3 long noncoding RNA regulates the TGF-beta pathway genes through formation of RNA-DNA triplex structures. *Nat Commun* **6**, 7743.
49 Stollar BD and Raso V (1974) Antibodies recognise specific structures of triple-helical polynucleotides built on poly(A) or poly(dA). *Nature* **250**, 231–234.
50 Schmitz KM, Mayer C, Postepska A and Grummt I (2010) Interaction of noncoding RNA with the rDNA promoter mediates recruitment of DNMT3b and silencing of rRNA genes. *Genes Dev* **24**, 2264–2269.
51 Bacolla A, Wang G and Vasquez KM (2015) New perspectives on DNA and RNA triplexes as effectors of biological activity. *PLoS Genet* **11**, e1005696.
52 Grummt I (2010) Wisely chosen paths – regulation of rRNA synthesis: delivered on 30 June at the 35th FEBS Congress in Gothenburg, Sweden. *FEBS J* **277**, 4626–4639.
53 Lafontaine DL (2015) Noncoding RNAs in eukaryotic ribosome biogenesis and function. *Nat Struct Mol Biol* **22**, 11–19.
54 Bierhoff H, Postepska-Igielska A and Grummt I (2013) Noisy silence: non-coding RNA and heterochromatin formation at repetitive elements. *Epigenetics* **9**, 53–61.
55 Strohner R, Nemeth A, Jansa P, Hofmann-Rohrer U, Santoro R, Langst G and Grummt I (2001) NoRC – a novel member of mammalian ISWI-containing chromatin remodeling machines. *EMBO J* **20**, 4892–4900.
56 Nemeth A, Strohner R, Grummt I and Langst G (2004) The chromatin remodeling complex NoRC and TTF-I cooperate in the regulation of the mammalian rRNA genes in vivo. *Nucleic Acids Res* **32**, 4091–4099.
57 Santoro R and Grummt I (2005) Epigenetic mechanism of rRNA gene silencing: temporal order of NoRC-mediated histone modification, chromatin remodeling, and DNA methylation. *Mol Cell Biol* **25**, 2539–2546.
58 Jeon Y and Lee JT (2011) YY1 tethers Xist RNA to the inactive X nucleation center. *Cell* **146**, 119–133.
59 Soruco MM, Chery J, Bishop EP, Siggers T, Tolstorukov MY, Leydon AR, Sugden AU, Goebel K, Feng J, Xia P *et al*. (2013) The CLAMP protein links the MSL complex to the X chromosome during *Drosophila* dosage compensation. *Genes Dev* **27**, 1551–1556.
60 Hacisuleyman E, Goff LA, Trapnell C, Williams A, Henao-Mejia J, Sun L, McClanahan P, Hendrickson DG, Sauvageau M, Kelley DR *et al*. (2014) Topological organization of multichromosomal regions by the long intergenic noncoding RNA Firre. *Nat Struct Mol Biol* **21**, 198–206.
61 Hasegawa Y, Brockdorff N, Kawano S, Tsutui K, Tsutui K and Nakagawa S (2010) The matrix protein hnRNP U is required for chromosomal localization of Xist RNA. *Dev Cell* **19**, 469–476.
62 Wang KC, Yang YW, Liu B, Sanyal A, Corces-Zimmerman R, Chen Y, Lajoie BR, Protacio A, Flynn RA, Gupta RA *et al*. (2011) A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature* **472**, 120–124.
63 Engreitz JM, Pandya-Jones A, McDonel P, Shishkin A, Sirokman K, Surka C, Kadri S, Xing J, Goren A, Lander ES *et al*. (2013) The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. *Science* **341**, 1237973.
64 Simon MD, Pinter SF, Fang R, Sarma K, Rutenberg-Schoenberg M, Bowman SK, Kesner BA, Maier VK, Kingston RE and Lee JT (2013) High-resolution Xist binding maps reveal two-step

spreading during X-chromosome inactivation. *Nature* **504**, 465–469.

65 Ramirez F, Lingg T, Toscano S, Lam KC, Georgiev P, Chung HR, Lajoie BR, de Wit E, Zhan Y, de Laat W *et al*. (2015) High-affinity sites form an interaction network to facilitate spreading of the MSL complex across the X chromosome in *Drosophila*. *Mol Cell* **60**, 146–162.

66 Guetg C, Lienemann P, Sirri V, Grummt I, Hernandez-Verdun D, Hottiger MO, Fussenegger M and Santoro R (2010) The NoRC complex mediates the heterochromatin formation and stability of silent rRNA genes and centromeric repeats. *EMBO J* **29**, 2135–2146.

67 Rinn J and Guttman M (2014) RNA function. RNA and dynamic nuclear organization. *Science* **345**, 1240–1241.

68 Li X, Wu Z, Fu X and Han W (2014) lncRNAs: insights into their function and mechanics in underlying disorders. *Mutat Res, Rev Mutat Res* **762**, 1–21.

69 Schmitz SU, Grote P and Herrmann BG (2016) Mechanisms of long noncoding RNA function in development and disease. *Cell Mol Life Sci* **73**, 2491–2509.

4. Discussion

In the present work we have shown that the chromatin state of rRNA genes, which are the genetic component of the nucleolus, has a crucial function in the process of ESC differentiation and is required to allow ESCs to exit pluripotency. During differentiation, the genome of ESCs undergoes massive remodeling, changing from an open euchromatic state to a more compact heterochromatic structure. This change in chromatin structure marks the exit from the pluripotent state and is linked to the acquisition of transcriptional programs that lead to specific lineage commitment and cell differentiation (Gaspar-Maia et al., 2011).

Together with the rest of the genome also part of rRNA genes experience the change in chromatin conformation and acquire heterochromatic features during differentiation (Savic et al., 2014). We have previously shown that formation of heterochromatin at the rDNA actually induces other parts of the genome to acquire a more compact structure (Savic et al., 2014). In particular, addition of mature lncRNA pRNA in ESCs was sufficient to specifically induce heterochromatin at rRNA genes as a consequence of TIP5 recruitment and this additionally resulted in the appearance of highly condensed regions at the periphery of the nucleolus. These results suggested that changes in the chromatin state of rRNA genes could affect the chromatin state of other regions of the genome, implying a broader influence of nucleolar chromatin on genome architecture. In fact, upon establishment of rDNA heterochromatin in ESCs, also other portions of the genome, and in particular repetitive sequences like major and minor satellites, experienced transcriptional repression acquiring heterochromatic marks like H3K9me2. The crosstalk between rDNA and pericentric heterochromatin is in line with previous results from our and other laboratories showing that TIP5 depletion in NIH3T3 cells caused not only the reduction of rRNA gene silencing but resulted also in the loss of perinucleolar heterochromatin and reduction of silent histone marks at pericentric heterochromatin (Guetg et al., 2010; Postepska-Igielska et al., 2013). Furthermore the link between rRNA genes and chromatin architecture of the rest of the genome is also supported by previous results in Drosophila showing that deletion of rRNA repeats reduced heterochromatin content elsewhere in the genome and the extent of the rDNA deletion correlates with the loss of silencing in

much the same manner as mutations in known protein heterochromatin components (Paredes and Maggert, 2009).

Although the mechanisms through which rDNA heterochromatin formation guides spreading of genome compaction is not yet clear, these results highlight a crucial role for the nucleolar compartment in shaping genome architecture.

We have previously shown that in ESCs pRNA is absent due to impairment of processing of its precursor IGS-rRNA. Induction of ESCs differentiation resulted in activation of IGS-rRNA processing and accumulation of mature pRNA, leading to recruitment of TIP5 to rDNA and formation of heterochromatin (Savic et al., 2014).

In the present work we aimed to elucidate the mechanism of pRNA maturation in order to better understand the importance of rDNA heterochromatin formation in the context of ESC differentiation. Our results identified the RNA helicase DHX9 as essential for IGS-rRNA processing into pRNA in differentiated cells. In differentiated cells, DHX9 is localized within nucleoli and associates with TIP5. Remarkably, depletion of DHX9 in differentiated cells resulted in TIP5 displacement from nucleoli, decrease of TIP5 binding to rDNA promoters and reduction of heterochromatic marks at rRNA genes, suggesting an essential role for DHX9 in the recruitment of TIP5 to rRNA genes and establishment of rDNA heterochromatin.

Depletion of DHX9 in ESCs impaired the ability to differentiate without affecting their self-renewal capacity. In particular, induction of differentiation of DHX9 depleted ESCs resulted in massive cells death implicating an essential role for DHX9 in the first phases of differentiation. This was in line with previous data showing the importance of DHX9 in early development. Indeed DHX9 KO mouse embryos showed impaired gastrulation and did not develop further than E7.5 (Lee et al., 1998). Strikingly we found that introduction of mature pRNA in DHX9 depleted ESCs was sufficient to allow differentiation. This demonstrates that DHX9-mediated maturation of pRNA is a key event that allows exit from pluripotency. Additionally our observation of the switch in DHX9 localization from a spread nuclear distribution to a confined nucleolar positioning during differentiation and its increased binding to rDNA well correlate with the activation of IGS-rRNA processing, which takes place in nucleoli only upon ESCs differentiation. Nevertheless, it is still unclear what determines the change of DHX9 localization during differentiation and it will be the aim of future studies.

The results described here depict a model in which DHX9 mediated IGS-rRNA processing represents the key step in the formation of rDNA heterochromatin and, in a larger scale, in the initiation of genome architecture remodeling that results from it. Although the set up of the described experiments does not allow us to directly distinguish whether impairment of ESC differentiation is due to up regulation of rRNA transcription or the lack of heterochromatin per se, we favor the latter case. Indeed, it is unlikely that cells with increased rRNA levels undergo cell death since ribosome biogenesis is well known to be positively correlated with cell viability and proliferation (Moss and Stefanovsky, 2002). Thus, our results suggest that the function of rRNA genes might not only be limited to the synthesis of rRNA. This model is in agreement with early studies showing that the fraction of silent rRNA genes present in each differentiated cell does not change its transcriptional state even under conditions of high metabolic activities, suggesting a role that is not related to the production of ribosomes (Conconi et al., 1989). This is also in line with our previous finding showing that the impairment of rDNA heterochromatin formation due to depletion of TIP5 in ESCs resulted in impairment of differentiation (Savic et al., 2014). Furthermore it was previously shown that reduction of rDNA heterochromatin in differentiated cells is linked with genomic instability in yeast, flies and mammals (Guetg et al., 2010; Peng and Karpen, 2007; Straight et al., 1999), suggesting that acquisition of rDNA heterochromatin during differentiation is essential to guarantee genome stability. Thus, our data favor a model in which the nucleolus is not only the cellular compartment where ribosomes are produced but it is also a central component of nuclear architecture that coordinates the balance between euchromatin and heterochromatin according to developmental stages. This is supported by other works showing that, during differentiation, repressed and heterochromatic portions of the genome contact either the nuclear lamina or the nucleoli (Peric-Hupkes et al., 2010).

Despite being necessary for IGS-rRNA processing, DHX9 is known to have only a catalytic domain with NTP hydrolytic activity enclosed between the two Rec-like domains and no nuclease activity has been reported to date (Bartova et al., 2008; Lee and Pelletier, 2016). Interestingly, our analysis of DHX9 interactome revealed a high enrichment for components of the spliceosome complex (i.e. U2 Small Nuclear RNA Auxiliary Factor 2 U2AF2, Splicing factor 3B subunit 3 SF3B3, pre-mRNA Processing Factor 3 and 8, PRPF3 and 8) suggesting that the splicing pathway might

be implicated in the processing of IGS-rRNA into pRNA. In fact the spliceosome was also the major hit obtained through cellular component and pathway analysis of IGS-rRNA pulled-down proteins from ESCs, differentiated ESCs and NIH3T3 nuclear extracts. The involvement of spliceosome in IGS-rRNA processing is also suggested by the absence of any endo- or exonucleolytic enzyme in the analysis of IGS-rRNA interacting proteins. Moreover, Xrn2 (5'-3' Exoribonuclease 2) was the unique ribonuclease we identified as DHX9-associated proteins in three out of four experiments. However, a careful analysis revealed that Xrn2 is not implicated in IGS-rRNA processing (data not shown). Thus our favorite hypothesis is that DHX9 recruits or is part of a specific subtype of spliceosome complex involved in IGS-rRNA processing and that this complex can assembly on IGS-rRNA only at the onset of ESC differentiation. The specificity of this complex is also supported by the absence, among the DHX9 and IGS-rRNA interacting factors, of the canonical DExH/RHA helicases of the spliceosome complex normally involved in pre-mRNA splicing (i.e. PRPF 2, 16, 22 and 43 (Jarmoskaite and Russell, 2014)).

The reason why DHX9 is absent in nucleoli and rRNA genes of ESCs still remains unknown and understanding this process will be aim of our future studies. It cannot be excluded at this point that posttranslational modifications of DHX9 or posttranscriptional modifications of IGS-rRNA could be essential to activate the processing. Very little is known about DHX9 PTMs. In particular two works have shown that DHX9 can be phosphorylated. In one case the N-terminal part of DHX9, corresponding to the dsRBD domain, can be phosphorylated by protein kinase R (PKR also known as EIF2AK2) as a response of the innate immune system to human immunodeficiency virus (HIV) infection. This modification results in the reduction of viral RNA binding by DHX9 and inhibits the capacity of DHX9 to enhance expression of genetic elements encoded by HIV (Sadler et al., 2009). DHX9 can also be phosphorylated by the DNA-dependent protein kinase (DNA-PK), a complex formed by the DNA binding protein Ku and the DNA-PK catalytic subunit (also known as XRCC7) that is involved in the non-homologous end-joining (NHEJ) pathway of DNA repair. Interestingly DHX9 phosphorylation by DNA-PK requires RNA and results in a different distribution of DHX9 in the nucleus (Zhang et al., 2004). Though this study was conducted on cervical cancer HeLa cells in which DHX9 is singularly excluded from nucleoli and DNA-PK activity on DHX9 was shown only *in vitro,* it would be worth to investigate the role of DHX9

phosphorylation as a possible mechanism at the basis of DHX9 differential localization in ESCs and differentiated cells. Furthermore a possible effect of DHX9 modification on its RNA binding ability, as suggested by the study on PKR, could also reinforce the possible regulation of the timing of IGS-rRNA processing due to the different DHX9 localization.

One of the most common RNA posttranscriptional modifications is the methylation of the nitrogen at the position 6 of the adenine purine ring (m^6A), catalyzed by the enzymes methyltransferase-like 3 and 14 (MTTL3 and 14). Knock out of MTTL3 is embryonic lethal because implantation of the embryo is impaired. However ESCs can be derived from early blastocysts of *Mettl3*$^{-/-}$ mice and these show normal morphology but impaired differentiation due to stabilization of pluripotency factors mRNAs that lead to a "hyperpluripotent" phenotype (Batista et al., 2014; Geula et al., 2015). A role for m^6A was previously demonstrated in the context of regulation of gene expression through mRNA splicing, localization, and degradation, as well as in modulating the RNA binding capacity of m^6A binding "reader" proteins (Dominissini et al., 2012). However all these works focus on the effect of m^6A on mRNAs and very little is known about this modification on the non-coding transcriptome especially in the context of developmental processes. A recent work described that m^6A is necessary for Xist mediated XCI during ESC differentiation (Patil et al., 2016). In particular in this work it was shown that depletion of adaptor proteins RBM15 and RBM15B that are responsible for METTL3 recruitment on Xist determine absence of adenine methylation and impairment of XCI. The role of m^6A in inducing XCI is mediated by a reader protein belonging to the YTH family that is able to specifically recognize m^6A on Xist and interact with repressor proteins involved in XCI like SHARP, LBR, HNRNPU and PRC2 components. This is the first example of functional characterization of m^6A posttranscriptional modification in lncRNAs and highlights a system of "writer" and "readers" definitely akin to the epigenetic one. This suggests a possible "epitranscriptome" that could contribute together with the epigenome in establishing further layers of regulation of gene expression by posttranscriptional modification of mRNA and lncRNA (Dominissini, 2014).

Despite we did not identify any of the classical m^6A reader proteins among the DHX9 or IGS-rRNA interacting factors we cannot exclude a role for this RNA modification in regulating the induction of IGS-rRNA processing during ESC differentiation. Thus further studies will be needed to address the existence of a

differential pattern of lncRNA posttranscriptional modification between ESCs and differentiated cells and the effects of these modifications.

Taken together, the results of our work revealed the importance of lncRNA processing in modulating chromatin structure. In particular, our work highlighted the cellular function of the nucleolar chromatin in seeding euchromatin and heterochromatin compartments as a key step in the maintenance of pluripotency and the initiation of differentiation processes.

Bibliography

Aalto, A.P., and Pasquinelli, A.E. (2012). Small non-coding RNAs mount a silent revolution in gene expression. Curr Opin Cell Biol *24*, 333-340.

Alva, V., Ammelburg, M., Soding, J., and Lupas, A.N. (2007). On the origin of the histone fold. BMC Struct Biol *7*, 17.

Andersen, J.S., Lyon, C.E., Fox, A.H., Leung, A.K., Lam, Y.W., Steen, H., Mann, M., and Lamond, A.I. (2002). Directed proteomic analysis of the human nucleolus. Curr Biol *12*, 1-11.

Aregger, R., and Klostermeier, D. (2009). The DEAD box helicase YxiN maintains a closed conformation during ATP hydrolysis. Biochemistry *48*, 10679-10681.

Avilion, A.A., Nicolis, S.K., Pevny, L.H., Perez, L., Vivian, N., and Lovell-Badge, R. (2003). Multipotent cell lineages in early mouse development depend on SOX2 function. Genes Dev *17*, 126-140.

Azuara, V., Perry, P., Sauer, S., Spivakov, M., Jorgensen, H.F., John, R.M., Gouti, M., Casanova, M., Warnes, G., Merkenschlager, M.*, et al.* (2006). Chromatin signatures of pluripotent cell lines. Nat Cell Biol *8*, 532-538.

Bartel, D.P. (2004). MicroRNAs: genomics, biogenesis, mechanism, and function. Cell *116*, 281-297.

Bartova, E., Galiova, G., Krejci, J., Harnicarova, A., Strasak, L., and Kozubek, S. (2008). Epigenome and chromatin structure in human embryonic stem cells undergoing differentiation. Dev Dyn *237*, 3690-3702.

Basnet, H., Su, X.B., Tan, Y., Meisenhelder, J., Merkurjev, D., Ohgi, K.A., Hunter, T., Pillus, L., and Rosenfeld, M.G. (2014). Tyrosine phosphorylation of histone H2A by CK2 regulates transcriptional elongation. Nature *516*, 267-271.

Bassett, A.R., Akhtar, A., Barlow, D.P., Bird, A.P., Brockdorff, N., Duboule, D., Ephrussi, A., Ferguson-Smith, A.C., Gingeras, T.R., Haerty, W.*, et al.* (2014). Considerations when investigating lncRNA function in vivo. Elife *3*, e03058.

Batista, P.J., Molinie, B., Wang, J., Qu, K., Zhang, J., Li, L., Bouley, D.M., Lujan, E., Haddad, B., Daneshvar, K.*, et al.* (2014). m(6)A RNA modification controls cell fate transition in mammalian embryonic stem cells. Cell Stem Cell *15*, 707-719.

Baubec, T., Colombo, D.F., Wirbelauer, C., Schmidt, J., Burger, L., Krebs, A.R., Akalin, A., and Schubeler, D. (2015). Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation. Nature *520*, 243-247.

Bell, P., Dabauvalle, M.C., and Scheer, U. (1992). In vitro assembly of prenucleolar bodies in Xenopus egg extract. J Cell Biol *118*, 1297-1304.

Berger, C., Horlebein, A., Gogel, E., and Grummt, F. (1997). Temporal order of replication of mouse ribosomal RNA genes during the cell cycle. Chromosoma *106*, 479-484.

Bertone, P., Stolc, V., Royce, T.E., Rozowsky, J.S., Urban, A.E., Zhu, X., Rinn, J.L., Tongprasit, W., Samanta, M., Weissman, S.*, et al.* (2004). Global identification of human transcribed sequences with genome tiling arrays. Science *306*, 2242-2246.

Bestor, T.H. (2000). The DNA methyltransferases of mammals. Hum Mol Genet *9*, 2395-2402.

Bhaskaran, H., and Russell, R. (2007). Kinetic redistribution of native and misfolded RNAs by a DEAD-box chaperone. Nature *449*, 1014-1018.

Bhattacharya, D., Talwar, S., Mazumder, A., and Shivashankar, G.V. (2009). Spatio-temporal plasticity in chromatin organization in mouse cell differentiation and during Drosophila embryogenesis. Biophys J *96*, 3832-3839.

Bibel, M., Richter, J., Schrenk, K., Tucker, K.L., Staiger, V., Korte, M., Goetz, M., and Barde, Y.A. (2004). Differentiation of mouse embryonic stem cells into a defined neuronal lineage. Nat Neurosci *7*, 1003-1009.

Bierhoff, H., Postepska-Igielska, A., and Grummt, I. (2014). Noisy silence: non-coding RNA and heterochromatin formation at repetitive elements. Epigenetics *9*, 53-61.

Black, J.C., Van Rechem, C., and Whetstine, J.R. (2012). Histone lysine methylation dynamics: establishment, regulation, and biological impact. Mol Cell *48*, 491-507.

Bochar, D.A., Savard, J., Wang, W., Lafleur, D.W., Moore, P., Cote, J., and Shiekhattar, R. (2000). A family of chromatin remodeling factors related to Williams syndrome transcription factor. Proc Natl Acad Sci U S A *97*, 1038-1043.

Borgel, J., Guibert, S., Li, Y., Chiba, H., Schubeler, D., Sasaki, H., Forne, T., and Weber, M. (2010). Targets and dynamics of promoter DNA methylation during early mouse development. Nat Genet *42*, 1093-1100.

Borsani, G., Tonlorenzi, R., Simmler, M.C., Dandolo, L., Arnaud, D., Capra, V., Grompe, M., Pizzuti, A., Muzny, D., Lawrence, C.*, et al.* (1991). Characterization of a murine gene expressed from the inactive X chromosome. Nature *351*, 325-329.

Bourc'his, D., and Bestor, T.H. (2004). Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. Nature *431*, 96-99.

Bourgeois, C.F., Mortreux, F., and Auboeuf, D. (2016). The multiple functions of RNA helicases as drivers and regulators of gene expression. Nat Rev Mol Cell Biol *17*, 426-438.

Boyer, L.A., Lee, T.I., Cole, M.F., Johnstone, S.E., Levine, S.S., Zucker, J.P., Guenther, M.G., Kumar, R.M., Murray, H.L., Jenner, R.G.*, et al.* (2005). Core transcriptional regulatory circuitry in human embryonic stem cells. Cell *122*, 947-956.

Bozhenok, L., Wade, P.A., and Varga-Weisz, P. (2002). WSTF-ISWI chromatin remodeling complex targets heterochromatic replication foci. Embo J *21*, 2231-2241.

Brockdorff, N., Ashworth, A., Kay, G.F., McCabe, V.M., Norris, D.P., Cooper, P.J., Swift, S., and Rastan, S. (1992). The product of the mouse Xist gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus. Cell *71*, 515-526.

Brockdorff, N., and Turner, B.M. (2015). Dosage compensation in mammals. Cold Spring Harb Perspect Biol *7*, a019406.

Brown, C.J., Hendrich, B.D., Rupert, J.L., Lafreniere, R.G., Xing, Y., Lawrence, J., and Willard, H.F. (1992). The human XIST gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. Cell *71*, 527-542.

Burdon, T., Stracey, C., Chambers, I., Nichols, J., and Smith, A. (1999). Suppression of SHP-2 and ERK signalling promotes self-renewal of mouse embryonic stem cells. Dev Biol *210*, 30-43.

Calo, E., Flynn, R.A., Martin, L., Spitale, R.C., Chang, H.Y., and Wysocka, J. (2015). RNA helicase DDX21 coordinates transcription and ribosomal RNA processing. Nature *518*, 249-253.

Caretti, G., Schiltz, R.L., Dilworth, F.J., Di Padova, M., Zhao, P., Ogryzko, V., Fuller-Pace, F.V., Hoffman, E.P., Tapscott, S.J., and Sartorelli, V. (2006). The RNA helicases p68/p72 and the noncoding RNA SRA are coregulators of MyoD and skeletal muscle differentiation. Dev Cell *11*, 547-560.

Carninci, P., Kasukawa, T., Katayama, S., Gough, J., Frith, M.C., Maeda, N., Oyama, R., Ravasi, T., Lenhard, B., Wells, C.*, et al.* (2005). The transcriptional landscape of the mammalian genome. Science *309*, 1559-1563.

Caudy, A.A., and Pikaard, C.S. (2002). Xenopus ribosomal RNA gene intergenic spacer elements conferring transcriptional enhancement and nucleolar dominance-like competition in oocytes. J Biol Chem *277*, 31577-31584.

Cech, T.R., and Steitz, J.A. (2014). The noncoding RNA revolution-trashing old rules to forge new ones. Cell *157*, 77-94.

Chambers, I., Silva, J., Colby, D., Nichols, J., Nijmeijer, B., Robertson, M., Vrana, J., Jones, K., Grotewold, L., and Smith, A. (2007). Nanog safeguards pluripotency and mediates germline development. Nature *450*, 1230-1234.

Chen, C., Nott, T.J., Jin, J., and Pawson, T. (2011). Deciphering arginine methylation: Tudor tells the tale. Nat Rev Mol Cell Biol *12*, 629-642.

Chen, C.K., Blanco, M., Jackson, C., Aznauryan, E., Ollikainen, N., Surka, C., Chow, A., Cerase, A., McDonel, P., and Guttman, M. (2016). Xist recruits the X chromosome to the nuclear lamina to enable chromosome-wide silencing. Science.

Cheng, J., Kapranov, P., Drenkow, J., Dike, S., Brubaker, S., Patel, S., Long, J., Stern, D., Tammana, H., Helt, G.*, et al.* (2005). Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution. Science *308*, 1149-1154.

Chu, C., Zhang, Q.C., da Rocha, S.T., Flynn, R.A., Bharadwaj, M., Calabrese, J.M., Magnuson, T., Heard, E., and Chang, H.Y. (2015). Systematic discovery of Xist RNA binding proteins. Cell *161*, 404-416.

Chun, Y.S., Chaudhari, P., and Jang, Y.Y. (2010). Applications of patient-specific induced pluripotent stem cells; focused on disease modeling, drug screening and therapeutic potentials for liver disease. Int J Biol Sci *6*, 796-805.

Clerget, G., Abel, Y., and Rederstorff, M. (2015). Small non-coding RNAs: a quick look in the rearview mirror. Methods Mol Biol *1296*, 3-9.

Collins, K. (2006). The biogenesis and regulation of telomerase holoenzymes. Nat Rev Mol Cell Biol *7*, 484-494.

Conconi, A., Widmer, R.M., Koller, T., and Sogo, J.M. (1989). Two different chromatin structures coexist in ribosomal RNA genes throughout the cell cycle. Cell *57*, 753-761.

Crampton, N., Bonass, W.A., Kirkham, J., Rivetti, C., and Thomson, N.H. (2006). Collision events between RNA polymerases in convergent transcription studied by atomic force microscopy. Nucleic Acids Res *34*, 5416-5425.

Cuthbert, G.L., Daujat, S., Snowden, A.W., Erdjument-Bromage, H., Hagiwara, T., Yamada, M., Schneider, R., Gregory, P.D., Tempst, P., Bannister, A.J.*, et al.* (2004). Histone deimination antagonizes arginine methylation. Cell *118*, 545-553.

Czaplinski, K., Kocher, T., Schelder, M., Segref, A., Wilm, M., and Mattaj, I.W. (2005). Identification of 40LoVe, a Xenopus hnRNP D family protein involved in localizing a TGF-beta-related mRNA during oogenesis. Dev Cell *8*, 505-515.

De Winter, R.F., and Moss, T. (1987). A complex array of sequences enhances ribosomal transcription in Xenopus laevis. J Mol Biol *196*, 813-827.

Deaton, A.M., and Bird, A. (2011). CpG islands and the regulation of transcription. Genes Dev *25*, 1010-1022.

Dev, V.G., Tantravahi, R., Miller, D.A., and Miller, O.J. (1977). Nucleolus organizers in Mus musculus subspecies and in the RAG mouse cell line. Genetics *86*, 389-398.

Djebali, S., Davis, C.A., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, A., Tanzer, A., Lagarde, J., Lin, W., Schlesinger, F.*, et al.* (2012). Landscape of transcription in human cells. Nature *489*, 101-108.

Dominissini, D. (2014). Genomics and Proteomics. Roadmap to the epitranscriptome. Science *346*, 1192.

Dominissini, D., Moshitch-Moshkovitz, S., Schwartz, S., Salmon-Divon, M., Ungar, L., Osenberg, S., Cesarkas, K., Jacob-Hirsch, J., Amariglio, N., Kupiec, M.*, et al.* (2012). Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. Nature *485*, 201-206.

Dorigo, B., Schalch, T., Kulangara, A., Duda, S., Schroeder, R.R., and Richmond, T.J. (2004). Nucleosome arrays reveal the two-start organization of the chromatin fiber. Science *306*, 1571-1573.

Dundr, M., Misteli, T., and Olson, M.O. (2000). The dynamics of postmitotic reassembly of the nucleolus. J Cell Biol *150*, 433-446.

Efroni, S., Duttagupta, R., Cheng, J., Dehghani, H., Hoeppner, D.J., Dash, C., Bazett-Jones, D.P., Le Grice, S., McKay, R.D., Buetow, K.H.*, et al.* (2008). Global transcription in pluripotent embryonic stem cells. Cell Stem Cell *2*, 437-447.

Ehrlich, M., Gama-Sosa, M.A., Huang, L.H., Midgett, R.M., Kuo, K.C., McCune, R.A., and Gehrke, C. (1982). Amount and distribution of 5-methylcytosine in human DNA from different types of tissues of cells. Nucleic Acids Res *10*, 2709-2721.

Engreitz, J.M., Pandya-Jones, A., McDonel, P., Shishkin, A., Sirokman, K., Surka, C., Kadri, S., Xing, J., Goren, A., Lander, E.S.*, et al.* (2013). The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. Science *341*, 1237973.

Epsztejn-Litman, S., Feldman, N., Abu-Remaileh, M., Shufaro, Y., Gerson, A., Ueda, J., Deplus, R., Fuks, F., Shinkai, Y., Cedar, H.*, et al.* (2008). De novo DNA methylation promoted by G9a prevents reprogramming of embryonically silenced genes. Nat Struct Mol Biol *15*, 1176-1183.

Etheridge, K.T., Banik, S.S., Armbruster, B.N., Zhu, Y., Terns, R.M., Terns, M.P., and Counter, C.M. (2002). The nucleolar localization domain of the catalytic subunit of human telomerase. J Biol Chem *277*, 24764-24770.

Evans, M.J., and Kaufman, M.H. (1981). Establishment in culture of pluripotential cells from mouse embryos. Nature *292*, 154-156.

Fairman-Williams, M.E., Guenther, U.P., and Jankowsky, E. (2010). SF1 and SF2 helicases: family matters. Curr Opin Struct Biol *20*, 313-324.

Fatemi, M., and Wade, P.A. (2006). MBD family proteins: reading the epigenetic code. J Cell Sci *119*, 3033-3037.

Fatica, A., and Bozzoni, I. (2014). Long non-coding RNAs: new players in cell differentiation and development. Nat Rev Genet *15*, 7-21.

Felsenfeld, G., and Groudine, M. (2003). Controlling the double helix. Nature *421*, 448-453.

Fennessy, R.T., and Owen-Hughes, T. (2016). Establishment of a promoter-based chromatin architecture on recently replicated DNA can accommodate variable inter-nucleosome spacing. Nucleic Acids Res.

Finch, J.T., and Klug, A. (1976). Solenoidal model for superstructure in chromatin. Proc Natl Acad Sci U S A *73*, 1897-1901.

Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E., and Mello, C.C. (1998). Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. Nature *391*, 806-811.

Friedemann, J., Grosse, F., and Zhang, S. (2005). Nuclear DNA helicase II (RNA helicase A) interacts with Werner syndrome helicase and stimulates its exonuclease activity. J Biol Chem *280*, 31303-31313.

Fu, X.D., and Ares, M., Jr. (2014). Context-dependent control of alternative splicing by RNA-binding proteins. Nat Rev Genet *15*, 689-701.

Fuchsova, B., and Hozak, P. (2002). The localization of nuclear DNA helicase II in different nuclear compartments is linked to transcription. Exp Cell Res *279*, 260-270.

Fujii, R., Okamoto, M., Aratani, S., Oishi, T., Ohshima, T., Taira, K., Baba, M., Fukamizu, A., and Nakajima, T. (2001). A Role of RNA Helicase A in cis-Acting Transactivation Response Element-mediated Transcriptional Regulation of Human Immunodeficiency Virus Type 1. J Biol Chem *276*, 5445-5451.

Fuks, F., Hurd, P.J., Wolf, D., Nan, X., Bird, A.P., and Kouzarides, T. (2003). The methyl-CpG-binding protein MeCP2 links DNA methylation to histone methylation. J Biol Chem *278*, 4035-4040.

Fukunaga, A., Tanaka, A., and Oishi, K. (1975). Maleless, a recessive autosomal mutant of Drosophila melanogaster that specifically kills male zygotes. Genetics *81*, 135-141.

Fussner, E., Ahmed, K., Dehghani, H., Strauss, M., and Bazett-Jones, D.P. (2010). Changes in chromatin fiber density as a marker for pluripotency. Cold Spring Harb Symp Quant Biol *75*, 245-249.

Fussner, E., Strauss, M., Djuric, U., Li, R., Ahmed, K., Hart, M., Ellis, J., and Bazett-Jones, D.P. (2012). Open and closed domains in the mouse genome are configured as 10-nm chromatin fibres. EMBO Rep *13*, 992-996.

Gaspar-Maia, A., Alajem, A., Meshorer, E., and Ramalho-Santos, M. (2011). Open chromatin in pluripotency and reprogramming. Nat Rev Mol Cell Biol *12*, 36-47.

Gerber, J.K., Gogel, E., Berger, C., Wallisch, M., Muller, F., Grummt, I., and Grummt, F. (1997). Termination of mammalian rDNA replication: polar arrest of replication fork movement by transcription termination factor TTF-I. Cell *90*, 559-567.

Geula, S., Moshitch-Moshkovitz, S., Dominissini, D., Mansour, A.A., Kol, N., Salmon-Divon, M., Hershkovitz, V., Peer, E., Mor, N., Manor, Y.S.*, et al.* (2015). Stem cells. m6A mRNA methylation facilitates resolution of naive pluripotency toward differentiation. Science *347*, 1002-1006.

Gifford, C.A., Ziller, M.J., Gu, H., Trapnell, C., Donaghey, J., Tsankov, A., Shalek, A.K., Kelley, D.R., Shishkin, A.A., Issner, R.*, et al.* (2013). Transcriptional and epigenetic dynamics during specification of human embryonic stem cells. Cell *153*, 1149-1163.

Glatt, S., Alfieri, C., and Muller, C.W. (2011). Recognizing and remodeling the nucleosome. Curr Opin Struct Biol *21*, 335-341.

Gonzalez, I.L., and Sylvester, J.E. (1995). Complete sequence of the 43-kb human ribosomal DNA repeat: analysis of the intergenic spacer. Genomics *27*, 320-328.

Gorkin, D.U., Leung, D., and Ren, B. (2014). The 3D genome in transcriptional regulation and pluripotency. Cell Stem Cell *14*, 762-775.

Grimaldi, G., and Di Nocera, P.P. (1988). Multiple repeated units in Drosophila melanogaster ribosomal DNA spacer stimulate rRNA precursor transcription. Proc Natl Acad Sci U S A *85*, 5502-5506.

Grossniklaus, U., and Paro, R. (2014). Transcriptional silencing by polycomb-group proteins. Cold Spring Harb Perspect Biol *6*, a019331.

Grozdanov, P., Georgiev, O., and Karagyozov, L. (2003). Complete sequence of the 45-kb mouse ribosomal DNA repeat: analysis of the intergenic spacer. Genomics *82*, 637-643.

Grummt, I., Rosenbauer, H., Niedermeyer, I., Maier, U., and Ohrlein, A. (1986). A repeated 18 bp sequence motif in the mouse rDNA spacer mediates binding of a nuclear factor and transcription termination. Cell *45*, 837-846.

Guenther, M.G., Levine, S.S., Boyer, L.A., Jaenisch, R., and Young, R.A. (2007). A chromatin landmark and transcription initiation at most promoters in human cells. Cell *130*, 77-88.

Guetg, C., Lienemann, P., Sirri, V., Grummt, I., Hernandez-Verdun, D., Hottiger, M.O., Fussenegger, M., and Santoro, R. (2010). The NoRC complex mediates the heterochromatin formation and stability of silent rRNA genes and centromeric repeats. Embo J *29*, 2135-2146.

Guetg, C., Scheifele, F., Rosenthal, F., Hottiger, M.O., and Santoro, R. (2012). Inheritance of silent rDNA chromatin is mediated by PARP1 via noncoding RNA. Mol Cell *45*, 790-800.

Guil, S., and Esteller, M. (2012). Cis-acting noncoding RNAs: friends and foes. Nat Struct Mol Biol *19*, 1068-1075.

Guo, G., Wang, W., and Bradley, A. (2004). Mismatch repair genes identified using genetic screens in Blm-deficient embryonic stem cells. Nature *429*, 891-895.

Guttman, M., and Rinn, J.L. (2012). Modular regulatory principles of large non-coding RNAs. Nature *482*, 339-346.

Hagiwara, T., Nakashima, K., Hirano, H., Senshu, T., and Yamada, M. (2002). Deimination of arginine residues in nucleophosmin/B23 and histones in HL-60 granulocytes. Biochem Biophys Res Commun *290*, 979-983.

Hahn, M.W., and Wray, G.A. (2002). The g-value paradox. Evol Dev *4*, 73-75.

Haltiner, M.M., Smale, S.T., and Tjian, R. (1986). Two distinct promoter elements in the human rRNA gene identified by linker scanning mutagenesis. Mol Cell Biol *6*, 227-235.

Hartman, T.R., Qian, S., Bolinger, C., Fernandez, S., Schoenberg, D.R., and Boris-Lawrie, K. (2006). RNA helicase A is necessary for translation of selected messenger RNAs. Nat Struct Mol Biol *13*, 509-516.

Hartmuth, K., Urlaub, H., Vornlocher, H.P., Will, C.L., Gentzel, M., Wilm, M., and Luhrmann, R. (2002). Protein composition of human prespliceosomes isolated by a tobramycin affinity-selection method. Proc Natl Acad Sci U S A *99*, 16719-16724.

Hasegawa, Y., Brockdorff, N., Kawano, S., Tsutui, K., and Nakagawa, S. (2010). The matrix protein hnRNP U is required for chromosomal localization of Xist RNA. Dev Cell *19*, 469-476.

Hassa, P.O., Haenni, S.S., Elser, M., and Hottiger, M.O. (2006). Nuclear ADP-ribosylation reactions in mammalian cells: where are we today and where are we going? Microbiol Mol Biol Rev *70*, 789-829.

Hayashi, K., Ohta, H., Kurimoto, K., Aramaki, S., and Saitou, M. (2011). Reconstitution of the mouse germ cell specification pathway in culture by pluripotent stem cells. Cell *146*, 519-532.

He, Y.F., Li, B.Z., Li, Z., Liu, P., Wang, Y., Tang, Q., Ding, J., Jia, Y., Chen, Z., Li, L.*, et al.* (2011). Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. Science *333*, 1303-1307.

Henderson, A.S., Warburton, D., and Atwood, K.C. (1972). Location of ribosomal DNA in the human chromosome complement. Proc Natl Acad Sci U S A *69*, 3394-3398.

Hermann, A., Goyal, R., and Jeltsch, A. (2004). The Dnmt1 DNA-(cytosine-C5)-methyltransferase methylates DNA processively with high preference for hemimethylated target sites. J Biol Chem *279*, 48350-48359.

Huang, W., Thomas, B., Flynn, R.A., Gavzy, S.J., Wu, L., Kim, S.V., Hall, J.A., Miraldi, E.R., Ng, C.P., Rigo, F.*, et al.* (2015). DDX5 and its associated lncRNA Rmrp modulate TH17 cell effector functions. Nature *528*, 517-522.

Huarte, M., Guttman, M., Feldser, D., Garber, M., Koziol, M.J., Kenzelmann-Broz, D., Khalil, A.M., Zuk, O., Amit, I., Rabani, M.*, et al.* (2010). A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. Cell *142*, 409-419.

Hung, T., Wang, Y., Lin, M.F., Koegel, A.K., Kotake, Y., Grant, G.D., Horlings, H.M., Shah, N., Umbricht, C., Wang, P.*, et al.* (2011). Extensive and coordinated transcription of noncoding RNAs within cell-cycle promoters. Nat Genet *43*, 621-629.

Hutvagner, G., McLachlan, J., Pasquinelli, A.E., Balint, E., Tuschl, T., and Zamore, P.D. (2001). A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. Science *293*, 834-838.

Huynh, J.L., and Casaccia, P. (2013). Epigenetic mechanisms in multiple sclerosis: implications for pathogenesis and treatment. Lancet Neurol *12*, 195-206.

Ilik, I.A., Quinn, J.J., Georgiev, P., Tavares-Cadete, F., Maticzka, D., Toscano, S., Wan, Y., Spitale, R.C., Luscombe, N., Backofen, R.*, et al.* (2013). Tandem stem-loops in roX RNAs act together to mediate X chromosome dosage compensation in Drosophila. Mol Cell *51*, 156-173.

Inoue, A., and Zhang, Y. (2011). Replication-dependent loss of 5-hydroxymethylcytosine in mouse preimplantation embryos. Science *334*, 194.

Ito, T., Levenstein, M.E., Fyodorov, D.V., Kutach, A.K., Kobayashi, R., and Kadonaga, J.T. (1999). ACF consists of two subunits, Acf1 and ISWI, that function cooperatively in the ATP-dependent catalysis of chromatin assembly. Genes Dev *13*, 1529-1539.

Iwasaki, Y.W., Murano, K., Ishizu, H., Shibuya, A., Iyoda, Y., Siomi, M.C., Siomi, H., and Saito, K. (2016). Piwi Modulates Chromatin Accessibility by Regulating Multiple Factors Including Histone H1 to Repress Transposons. Mol Cell.

Jankowsky, E. (2011). RNA helicases at work: binding and rearranging. Trends Biochem Sci *36*, 19-29.

Jankowsky, E., and Bowers, H. (2006). Remodeling of ribonucleoprotein complexes with DExH/D RNA helicases. Nucleic Acids Res *34*, 4181-4188.

Jankowsky, E., and Fairman, M.E. (2007). RNA helicases--one fold for many functions. Curr Opin Struct Biol *17*, 316-324.

Jarmoskaite, I., and Russell, R. (2014). RNA helicase proteins as chaperones and remodelers. Annu Rev Biochem *83*, 697-725.

Jenuwein, T. (2006). The epigenetic magic of histone lysine methylation. Febs J *273*, 3121-3135.

Jeon, Y., and Lee, J.T. (2011). YY1 tethers Xist RNA to the inactive X nucleation center. Cell *146*, 119-133.

Jin, B., Tao, Q., Peng, J., Soo, H.M., Wu, W., Ying, J., Fields, C.R., Delmas, A.L., Liu, X., Qiu, J.*, et al.* (2008). DNA methyltransferase 3B (DNMT3B) mutations in ICF syndrome lead to altered epigenetic modifications and aberrant expression of genes regulating development, neurogenesis and immune function. Hum Mol Genet *17*, 690-709.

Jones, M.H., Hamana, N., Nezu, J., and Shimane, M. (2000). A novel family of bromodomain genes. Genomics *63*, 40-45.

Jones, P.A. (1999). The DNA methylation paradox. Trends Genet *15*, 34-37.

Jones, P.A. (2012). Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat Rev Genet *13*, 484-492.

Judson, R.L., Babiarz, J.E., Venere, M., and Blelloch, R. (2009). Embryonic stem cell-specific microRNAs promote induced pluripotency. Nat Biotechnol *27*, 459-461.

Jung, C., Mittler, G., Oswald, F., and Borggrefe, T. (2013). RNA helicase Ddx5 and the noncoding RNA SRA act as coactivators in the Notch signaling pathway. Biochim Biophys Acta *1833*, 1180-1189.

Kim, M., Trinh, B.N., Long, T.I., Oghamian, S., and Laird, P.W. (2004). Dnmt1 deficiency leads to enhanced microsatellite instability in mouse embryonic stem cells. Nucleic Acids Res *32*, 5742-5749.

Kim, V.N., Han, J., and Siomi, M.C. (2009). Biogenesis of small RNAs in animals. Nat Rev Mol Cell Biol *10*, 126-139.

Kino, T., Hurt, D.E., Ichijo, T., Nader, N., and Chrousos, G.P. (2010). Noncoding RNA gas5 is a growth arrest- and starvation-associated repressor of the glucocorticoid receptor. Sci Signal *3*, ra8.

Kiss, T. (2001). Small nucleolar RNA-guided post-transcriptional modification of cellular RNAs. Embo J *20*, 3617-3622.

Kohli, R.M., and Zhang, Y. (2013). TET enzymes, TDG and the dynamics of DNA demethylation. Nature *502*, 472-479.

Kouzarides, T. (2007). Chromatin modifications and their function. Cell *128*, 693-705.

Kretz, M., Siprashvili, Z., Chu, C., Webster, D.E., Zehnder, A., Qu, K., Lee, C.S., Flockhart, R.J., Groff, A.F., Chow, J.*, et al.* (2013). Control of somatic tissue differentiation by the long non-coding RNA TINCR. Nature *493*, 231-235.

Krishnan, J., and Mishra, R.K. (2014). Emerging trends of long non-coding RNAs in gene activation. Febs J *281*, 34-45.

Krol, J., Krol, I., Alvarez, C.P., Fiscella, M., Hierlemann, A., Roska, B., and Filipowicz, W. (2015). A network comprising short and long noncoding RNAs and RNA helicase controls mouse retina architecture. Nat Commun *6*, 7305.

Kuhn, A., Bartsch, I., and Grummt, I. (1990). Specific interaction of the murine transcription termination factor TTF I with class-I RNA polymerases. Nature *344*, 559-562.

Kuhn, A., and Grummt, I. (1987). A novel promoter in the mouse rDNA spacer is active in vivo and in vitro. Embo J *6*, 3487-3492.

Kuhn, A., and Grummt, I. (1992). Dual role of the nucleolar transcription factor UBF: trans-activator and antirepressor. Proc Natl Acad Sci U S A *89*, 7340-7344.

Kung, J.T., Colognori, D., and Lee, J.T. (2013). Long noncoding RNAs: past, present, and future. Genetics *193*, 651-669.

Langst, G., Becker, P.B., and Grummt, I. (1998). TTF-I determines the chromatin architecture of the active rDNA promoter. Embo J *17*, 3135-3145.

Langst, G., Blank, T.A., Becker, P.B., and Grummt, I. (1997). RNA polymerase I transcription on nucleosomal templates: the transcription termination factor TTF-I induces chromatin remodeling and relieves transcriptional repression. Embo J *16*, 760-768.

Learned, R.M., Learned, T.K., Haltiner, M.M., and Tjian, R.T. (1986). Human rRNA transcription is modulated by the coordinate binding of two factors to an upstream control element. Cell *45*, 847-857.

Lee, C.G., da Costa Soares, V., Newberger, C., Manova, K., Lacy, E., and Hurwitz, J. (1998). RNA helicase A is essential for normal gastrulation. Proc Natl Acad Sci U S A *95*, 13709-13713.

Lee, J.T. (2012). Epigenetic regulation by long noncoding RNAs. Science *338*, 1435-1439.

Lee, S., Kopp, F., Chang, T.C., Sataluri, A., Chen, B., Sivakumar, S., Yu, H., Xie, Y., and Mendell, J.T. (2016a). Noncoding RNA NORAD Regulates Genomic Stability by Sequestering PUMILIO Proteins. Cell *164*, 69-80.

Lee, T., Paquet, M., Larsson, O., and Pelletier, J. (2016b). Tumor cell survival dependence on the DHX9 DExH-box helicase. Oncogene.

Lee, T., and Pelletier, J. (2016). The biology of DHX9 and its potential as a therapeutic target. Oncotarget.

Leone, S., and Santoro, R. (2016). Challenges in the analysis of long noncoding RNA functionality. FEBS Lett *590*, 2342-2353.

LeRoy, G., Loyola, A., Lane, W.S., and Reinberg, D. (2000). Purification and characterization of a human factor that assembles and remodels chromatin. J Biol Chem *275*, 14787-14790.

Li, E., Beard, C., and Jaenisch, R. (1993). Role for DNA methylation in genomic imprinting. Nature *366*, 362-365.

Li, E., and Zhang, Y. (2014). DNA methylation in mammals. Cold Spring Harb Perspect Biol *6*, a019133.

Li, J., Langst, G., and Grummt, I. (2006). NoRC-dependent nucleosome positioning silences rRNA genes. Embo J *25*, 5735-5741.

Li, J., Santoro, R., Koberna, K., and Grummt, I. (2005). The chromatin remodeling complex NoRC controls replication timing of rRNA genes. Embo J *24*, 120-127.

Liang, G., and Zhang, Y. (2013). Embryonic stem cell and induced pluripotent stem cell: an epigenetic perspective. Cell Res *23*, 49-69.

Liao, J., Karnik, R., Gu, H., Ziller, M.J., Clement, K., Tsankov, A.M., Akopian, V., Gifford, C.A., Donaghey, J., Galonska, C.*, et al.* (2015). Targeted disruption of DNMT1, DNMT3A and DNMT3B in human embryonic stem cells. Nat Genet *47*, 469-478.

Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O.*, et al.* (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science *326*, 289-293.

Lin, C.H., Jackson, A.L., Guo, J., Linsley, P.S., and Eisenman, R.N. (2009). Myc-regulated microRNAs attenuate embryonic stem cell differentiation. Embo J *28*, 3157-3170.

Lin, L., Li, Y., Pyo, H.M., Lu, X., Raman, S.N., Liu, Q., Brown, E.G., and Zhou, Y. (2012). Identification of RNA helicase A as a cellular factor that interacts with influenza A virus NS1 protein and its role in the virus life cycle. J Virol *86*, 1942-1954.

Liu, F., Putnam, A., and Jankowsky, E. (2008). ATP hydrolysis is required for DEAD-box protein recycling but not for duplex unwinding. Proc Natl Acad Sci U S A *105*, 20209-20214.

Loh, Y.H., Wu, Q., Chew, J.L., Vega, V.B., Zhang, W., Chen, X., Bourque, G., George, J., Leong, B., Liu, J.*, et al.* (2006). The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. Nat Genet *38*, 431-440.

Long, E.O., and Dawid, I.B. (1980). Repeated genes in eukaryotes. Annu Rev Biochem *49*, 727-764.

Lucchesi, J.C., and Kuroda, M.I. (2015). Dosage compensation in Drosophila. Cold Spring Harb Perspect Biol *7*.

Luense, L.J., Wang, X., Schon, S.B., Weller, A.H., Lin Shiao, E., Bryant, J.M., Bartolomei, M.S., Coutifaris, C., Garcia, B.A., and Berger, S.L. (2016). Comprehensive analysis of histone post-translational modifications in mouse and human male germ cells. Epigenetics Chromatin *9*, 24.

Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997). Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature *389*, 251-260.

Ma, L., Bajic, V.B., and Zhang, Z. (2013). On the classification of long non-coding RNAs. RNA Biol *10*, 925-933.

Maenner, S., Muller, M., Frohlich, J., Langer, D., and Becker, P.B. (2013). ATP-dependent roX RNA remodeling by the helicase maleless enables specific association of MSL proteins. Mol Cell *51*, 174-184.

Maeshima, K., Imai, R., Hikima, T., and Joti, Y. (2014). Chromatin structure revealed by X-ray scattering analysis and computational modeling. Methods *70*, 154-161.

Maeshima, K., Rogge, R., Tamura, S., Joti, Y., Hikima, T., Szerlong, H., Krause, C., Herman, J., Seidel, E., DeLuca, J.*, et al.* (2016). Nucleosomal arrays self-assemble into supramolecular globular structures lacking 30-nm fibers. Embo J *35*, 1115-1132.

Malik, N., and Rao, M.S. (2013). A review of the methods for human iPSC derivation. Methods Mol Biol *997*, 23-33.

Marmorstein, R., and Zhou, M.M. (2014). Writers and readers of histone acetylation: structure, mechanism, and inhibition. Cold Spring Harb Perspect Biol *6*, a018762.

Martin, G.R. (1981). Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. Proc Natl Acad Sci U S A *78*, 7634-7638.

Marx, V. (2012). Epigenetics: Reading the second genomic code. Nature *491*, 143-147.

Masui, S., Nakatake, Y., Toyooka, Y., Shimosato, D., Yagi, R., Takahashi, K., Okochi, H., Okuda, A., Matoba, R., Sharov, A.A.*, et al.* (2007). Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells. Nat Cell Biol *9*, 625-635.

Matera, A.G., Terns, R.M., and Terns, M.P. (2007). Non-coding RNAs: lessons from the small nuclear and small nucleolar RNAs. Nat Rev Mol Cell Biol *8*, 209-220.

Matsui, Y., Zsebo, K., and Hogan, B.L. (1992). Derivation of pluripotential embryonic stem cells from murine primordial germ cells in culture. Cell *70*, 841-847.

Mayer, C., Neubert, M., and Grummt, I. (2008). The structure of NoRC-associated RNA is crucial for targeting the chromatin remodelling complex NoRC to the nucleolus. EMBO Rep *9*, 774-780.

Mayer, C., Schmitz, K.M., Li, J., Grummt, I., and Santoro, R. (2006). Intergenic transcripts regulate the epigenetic state of rRNA genes. Mol Cell *22*, 351-361.

McHugh, C.A., Chen, C.K., Chow, A., Surka, C.F., Tran, C., McDonel, P., Pandya-Jones, A., Blanco, M., Burghard, C., Moradian, A*., et al.* (2015). The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. Nature *521*, 232-236.

Meissner, A. (2010). Epigenetic modifications in pluripotent and differentiated cells. Nat Biotechnol *28*, 1079-1088.

Meller, V.H., and Rattner, B.P. (2002). The roX genes encode redundant male-specific lethal transcripts required for targeting of the MSL complex. Embo J *21*, 1084-1091.

Meshorer, E., and Misteli, T. (2006). Chromatin in pluripotent embryonic stem cells and differentiation. Nat Rev Mol Cell Biol *7*, 540-546.

Miller, O.L., Jr., and Beatty, B.R. (1969). Visualization of nucleolar genes. Science *164*, 955-957.

Mills, J.R., Malina, A., Lee, T., Di Paola, D., Larsson, O., Miething, C., Grosse, F., Tang, H., Zannis-Hadjopoulos, M., Lowe, S.W*., et al.* (2013). RNAi screening uncovers Dhx9 as a modifier of ABT-737 resistance in an Emu-myc/Bcl-2 mouse model. Blood *121*, 3402-3412.

Mischo, H.E., Hemmerich, P., Grosse, F., and Zhang, S. (2005). Actinomycin D induces histone gamma-H2AX foci and complex formation of gamma-H2AX with Ku70 and nuclear DNA helicase II. J Biol Chem *280*, 9586-9594.

Mitsui, K., Tokuzawa, Y., Itoh, H., Segawa, K., Murakami, M., Takahashi, K., Maruyama, M., Maeda, M., and Yamanaka, S. (2003). The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. Cell *113*, 631-642.

Mohammad, F., Mondal, T., Guseva, N., Pandey, G.K., and Kanduri, C. (2010). Kcnq1ot1 noncoding RNA mediates transcriptional gene silencing by interacting with Dnmt1. Development *137*, 2493-2499.

Mohn, F., and Schubeler, D. (2009). Genetics and epigenetics: stability and plasticity during cellular differentiation. Trends Genet *25*, 129-136.

Morales, Y., Caceres, T., May, K., and Hevel, J.M. (2016). Biochemistry and regulation of the protein arginine methyltransferases (PRMTs). Arch Biochem Biophys *590*, 138-152.

Moss, T., and Stefanovsky, V.Y. (2002). At the center of eukaryotic life. Cell *109*, 545-548.

Nagano, T., Mitchell, J.A., Sanz, L.A., Pauler, F.M., Ferguson-Smith, A.C., Feil, R., and Fraser, P. (2008). The Air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. Science *322*, 1717-1720.

Nakajima, T., Uchida, C., Anderson, S.F., Lee, C.G., Hurwitz, J., Parvin, J.D., and Montminy, M. (1997). RNA helicase A mediates association of CBP with RNA polymerase II. Cell *90*, 1107-1112.

Nan, X., Ng, H.H., Johnson, C.A., Laherty, C.D., Turner, B.M., Eisenman, R.N., and Bird, A. (1998). Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. Nature *393*, 386-389.

Nelson, C.J., Santos-Rosa, H., and Kouzarides, T. (2006). Proline isomerization of histone H3 regulates lysine methylation and gene expression. Cell *126*, 905-916.

Nemeth, A., Guibert, S., Tiwari, V.K., Ohlsson, R., and Langst, G. (2008). Epigenetic regulation of TTF-I-mediated promoter-terminator interactions of rRNA genes. Embo J *27*, 1255-1265.

Oh, Y., Wei, H., Ma, D., Sun, X., and Liew, R. (2012). Clinical applications of patient-specific induced pluripotent stem cells in cardiovascular medicine. Heart *98*, 443-449.

Okada, Y., Feng, Q., Lin, Y., Jiang, Q., Li, Y., Coffield, V.M., Su, L., Xu, G., and Zhang, Y. (2005). hDOT1L links histone methylation to leukemogenesis. Cell *121*, 167-178.

Okano, M., Bell, D.W., Haber, D.A., and Li, E. (1999). DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell *99*, 247-257.

Olins, A.L., and Olins, D.E. (1974). Spheroid chromatin units (v bodies). Science *183*, 330-332.

Olson, M.O., Dundr, M., and Szebeni, A. (2000). The nucleolus: an old factory with unexpected capabilities. Trends Cell Biol *10*, 189-196.

Ooi, S.K., and Bestor, T.H. (2008). The colorful history of active DNA demethylation. Cell *133*, 1145-1148.

Paalman, M.H., Henderson, S.L., and Sollner-Webb, B. (1995). Stimulation of the mouse rRNA gene promoter by a distal spacer promoter. Mol Cell Biol *15*, 4648-4656.

Pandey, R.R., Mondal, T., Mohammad, F., Enroth, S., Redrup, L., Komorowski, J., Nagano, T., Mancini-Dinardo, D., and Kanduri, C. (2008). Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. Mol Cell *32*, 232-246.

Panov, K.I., Friedrich, J.K., Russell, J., and Zomerdijk, J.C. (2006). UBF activates RNA polymerase I transcription by stimulating promoter escape. Embo J *25*, 3310-3322.

Paredes, S., and Maggert, K.A. (2009). Ribosomal DNA contributes to global chromatin regulation. Proc Natl Acad Sci U S A *106*, 17829-17834.

Patil, D.P., Chen, C.K., Pickering, B.F., Chow, A., Jackson, C., Guttman, M., and Jaffrey, S.R. (2016). m6A RNA methylation promotes XIST-mediated transcriptional repression. Nature *537*, 369-373.

Peng, J.C., and Karpen, G.H. (2007). H3K9 methylation and RNA interference regulate nucleolar organization and repeated DNA stability. Nat Cell Biol *9*, 25-35.

Peric-Hupkes, D., Meuleman, W., Pagie, L., Bruggeman, S.W., Solovei, I., Brugman, W., Graf, S., Flicek, P., Kerkhoven, R.M., van Lohuizen, M., *et al.* (2010). Molecular maps of the reorganization of genome-nuclear lamina interactions during differentiation. Mol Cell *38*, 603-613.

Pertea, M., and Salzberg, S.L. (2010). Between a chicken and a grape: estimating the number of human genes. Genome Biol *11*, 206.

Pikaard, C.S., Pape, L.K., Henderson, S.L., Ryan, K., Paalman, M.H., Lopata, M.A., Reeder, R.H., and Sollner-Webb, B. (1990). Enhancers for RNA polymerase I in mouse ribosomal DNA. Mol Cell Biol *10*, 4816-4825.

Politz, J.C., Yarovoi, S., Kilroy, S.M., Gowda, K., Zwieb, C., and Pederson, T. (2000). Signal recognition particle components in the nucleolus. Proc Natl Acad Sci U S A *97*, 55-60.

Postepska-Igielska, A., Krunic, D., Schmitt, N., Greulich-Bode, K.M., Boukamp, P., and Grummt, I. (2013). The chromatin remodelling complex NoRC safeguards genome stability by heterochromatin formation at telomeres and centromeres. EMBO Rep *14*, 704-710.

Quinn, J.J., and Chang, H.Y. (2016). Unique features of long non-coding RNA biogenesis and function. Nat Rev Genet *17*, 47-62.

Quinodoz, S., and Guttman, M. (2014). Long noncoding RNAs: an emerging link between gene regulation and nuclear organization. Trends Cell Biol *24*, 651-663.

Ramakrishnan, V. (1997). Histone structure and the organization of the nucleosome. Annu Rev Biophys Biomol Struct *26*, 83-112.

Razin, S.V., and Gavrilov, A.A. (2014). Chromatin without the 30-nm fiber: constrained disorder instead of hierarchical folding. Epigenetics *9*, 653-657.

Reik, W. (2007). Stability and flexibility of epigenetic gene regulation in mammalian development. Nature *447*, 425-432.

Rich, A., and RajBhandary, U.L. (1976). Transfer RNA: molecular structure, sequence, and properties. Annu Rev Biochem *45*, 805-860.

Richmond, T.J., and Davey, C.A. (2003). The structure of DNA in the nucleosome core. Nature *423*, 145-150.

Rinn, J.L., Kertesz, M., Wang, J.K., Squazzo, S.L., Xu, X., Brugmann, S.A., Goodnough, L.H., Helms, J.A., Farnham, P.J., Segal, E.*, et al.* (2007). Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. Cell *129*, 1311-1323.

Robb, G.B., and Rana, T.M. (2007). RNA helicase A interacts with RISC in human cells and functions in RISC loading. Mol Cell *26*, 523-537.

Robinson, P.J., Fairall, L., Huynh, V.A., and Rhodes, D. (2006). EM measurements define the dimensions of the "30-nm" chromatin fiber: evidence for a compact, interdigitated structure. Proc Natl Acad Sci U S A *103*, 6506-6511.

Rosner, M.H., Vigano, M.A., Ozato, K., Timmons, P.M., Poirier, F., Rigby, P.W., and Staudt, L.M. (1990). A POU-domain transcription factor in early stem cells and germ cells of the mammalian embryo. Nature *345*, 686-692.

Rothbart, S.B., and Strahl, B.D. (2014). Interpreting the language of histone and DNA modifications. Biochim Biophys Acta *1839*, 627-643.

Rousseaux, S., and Khochbin, S. (2015). Histone Acylation beyond Acetylation: Terra Incognita in Chromatin Biology. Cell J *17*, 1-6.

Russo, V.E., Martienssen, R.A., and Riggs, A.D. (1996). Epigenetic mechanisms of gene regulation (Cold Spring Harbor Laboratory Press).

Sadler, A.J., Latchoumanin, O., Hawkes, D., Mak, J., and Williams, B.R. (2009). An antiviral response directed by PKR phosphorylation of the RNA helicase A. PLoS Pathog *5*, e1000311.

Sakabe, K., Wang, Z., and Hart, G.W. (2010). Beta-N-acetylglucosamine (O-GlcNAc) is part of the histone code. Proc Natl Acad Sci U S A *107*, 19915-19920.

Santoro, R. (2005). The silence of the ribosomal RNA genes. Cell Mol Life Sci *62*, 2067-2079.

Santoro, R. (2011). The Epigenetics of the Nucleolus: Structure and Function of Active and Silent Ribosomal RNA Genes. Protein Rev *15*, 57-82.

Santoro, R. (2014). Analysis of chromatin composition of repetitive sequences: the ChIP-Chop assay. Methods Mol Biol *1094*, 319-328.

Santoro, R., and Grummt, I. (2001). Molecular mechanisms mediating methylation-dependent silencing of ribosomal gene transcription. Mol Cell *8*, 719-725.

Santoro, R., and Grummt, I. (2005). Epigenetic mechanism of rRNA gene silencing: temporal order of NoRC-mediated histone modification, chromatin remodeling, and DNA methylation. Mol Cell Biol *25*, 2539-2546.

Santoro, R., Li, J., and Grummt, I. (2002). The nucleolar remodeling complex NoRC mediates heterochromatin formation and silencing of ribosomal gene transcription. Nat Genet *32*, 393-396.

Santoro, R., Schmitz, K.M., Sandoval, J., and Grummt, I. (2010). Intergenic transcripts originating from a subclass of ribosomal DNA repeats silence ribosomal RNA genes in trans. EMBO Rep *11*, 52-58.

Sapiro, A.L., Deng, P., Zhang, R., and Li, J.B. (2015). Cis regulatory effects on A-to-I RNA editing in related Drosophila species. Cell Rep *11*, 697-703.

Sarma, K., Levasseur, P., Aristarkhov, A., and Lee, J.T. (2010). Locked nucleic acids (LNAs) reveal sequence requirements and kinetics of Xist RNA localization to the X chromosome. Proc Natl Acad Sci U S A *107*, 22196-22201.

Sarma, K., and Reinberg, D. (2005). Histone variants meet their match. Nat Rev Mol Cell Biol *6*, 139-149.

Savic, N., Bar, D., Leone, S., Frommel, S.C., Weber, F.A., Vollenweider, E., Ferrari, E., Ziegler, U., Kaech, A., Shakhova, O.*, et al.* (2014). lncRNA maturation to initiate heterochromatin formation in the nucleolus is required for exit from pluripotency in ESCs. Cell Stem Cell *15*, 720-734.

Schad, E., Tompa, P., and Hegyi, H. (2011). The relationship between proteome size, structural disorder and organism complexity. Genome Biol *12*, R120.

Schalch, T., Duda, S., Sargent, D.F., and Richmond, T.J. (2005). X-ray structure of a tetranucleosome and its implications for the chromatin fibre. Nature *436*, 138-141.

Scheer, U., and Hock, R. (1999). Structure and function of the nucleolus. Curr Opin Cell Biol *11*, 385-390.

Schlesinger, S., and Goff, S.P. (2015). Retroviral transcriptional regulation and embryonic stem cells: war and peace. Mol Cell Biol *35*, 770-777.

Schmitges, F.W., Prusty, A.B., Faty, M., Stutzer, A., Lingaraju, G.M., Aiwazian, J., Sack, R., Hess, D., Li, L., Zhou, S.*, et al.* (2011). Histone methylation by PRC2 is inhibited by active chromatin marks. Mol Cell *42*, 330-341.

Schmitz, K.M., Mayer, C., Postepska, A., and Grummt, I. (2010). Interaction of noncoding RNA with the rDNA promoter mediates recruitment of DNMT3b and silencing of rRNA genes. Genes Dev *24*, 2264-2269.

Scholer, H.R., Dressler, G.R., Balling, R., Rohdewohld, H., and Gruss, P. (1990). Oct-4: a germline-specific transcription factor mapping to the mouse t-complex. Embo J *9*, 2185-2195.

Schweet, R., and Heintz, R. (1966). Protein synthesis. Annu Rev Biochem *35*, 723-758.

Shah, M.Y., and Licht, J.D. (2011). DNMT3A mutations in acute myeloid leukemia. Nat Genet *43*, 289-290.

Shaw, P.J., and Jordan, E.G. (1995). The nucleolus. Annu Rev Cell Dev Biol *11*, 93-121.

Shipony, Z., Mukamel, Z., Cohen, N.M., Landan, G., Chomsky, E., Zeliger, S.R., Fried, Y.C., Ainbinder, E., Friedman, N., and Tanay, A. (2014). Dynamic and static maintenance of epigenetic memory in pluripotent and somatic cells. Nature *513*, 115-119.

Siomi, M.C., Sato, K., Pezic, D., and Aravin, A.A. (2011). PIWI-interacting small RNAs: the vanguard of genome defence. Nat Rev Mol Cell Biol *12*, 246-258.

Smith, A.G., Heath, J.K., Donaldson, D.D., Wong, G.G., Moreau, J., Stahl, M., and Rogers, D. (1988). Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides. Nature *336*, 688-690.

Smith, Z.D., Chan, M.M., Mikkelsen, T.S., Gu, H., Gnirke, A., Regev, A., and Meissner, A. (2012). A unique regulatory phase of DNA methylation in the early mammalian embryo. Nature *484*, 339-344.

Sogo, J.M., Ness, P.J., Widmer, R.M., Parish, R.W., and Koller, T. (1984). Psoralen-crosslinking of DNA as a probe for the structure of active nucleolar chromatin. J Mol Biol *178*, 897-919.

Stancheva, I., Lucchini, R., Koller, T., and Sogo, J.M. (1997). Chromatin structure and methylation of rat rRNA genes studied by formaldehyde fixation and psoralen cross-linking. Nucleic Acids Res *25*, 1727-1735.

Stefanovsky, V., Langlois, F., Gagnon-Kugler, T., Rothblum, L.I., and Moss, T. (2006). Growth factor signaling regulates elongation of RNA polymerase I transcription in mammals via UBF phosphorylation and r-chromatin remodeling. Mol Cell *21*, 629-639.

Strahl, B.D., and Allis, C.D. (2000). The language of covalent histone modifications. Nature *403*, 41-45.

Straight, A.F., Shou, W., Dowd, G.J., Turck, C.W., Deshaies, R.J., Johnson, A.D., and Moazed, D. (1999). Net1, a Sir2-associated nucleolar protein required for rDNA silencing and nucleolar integrity. Cell *97*, 245-256.

Straub, T., Zabel, A., Gilfillan, G.D., Feller, C., and Becker, P.B. (2013). Different chromatin interfaces of the Drosophila dosage compensation complex revealed by high-shear ChIP-seq. Genome Res *23*, 473-485.

Strohner, R., Nemeth, A., Jansa, P., Hofmann-Rohrer, U., Santoro, R., Langst, G., and Grummt, I. (2001). NoRC--a novel member of mammalian ISWI-containing chromatin remodeling machines. Embo J *20*, 4892-4900.

Sun, N., Longaker, M.T., and Wu, J.C. (2010). Human iPS cell-based therapy: considerations before clinical applications. Cell Cycle *9*, 880-885.

Taft, R.J., Pheasant, M., and Mattick, J.S. (2007). The relationship between non-protein-coding DNA and eukaryotic complexity. Bioessays *29*, 288-299.

Takahashi, K., and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell *126*, 663-676.

Tessarz, P., and Kouzarides, T. (2014). Histone core modifications regulating nucleosome structure and dynamics. Nat Rev Mol Cell Biol *15*, 703-708.

Thoma, F., Koller, T., and Klug, A. (1979). Involvement of histone H1 in the organization of the nucleosome and of the salt-dependent superstructures of chromatin. J Cell Biol *83*, 403-427.

Tsai, M.C., Manor, O., Wan, Y., Mosammaparast, N., Wang, J.K., Lan, F., Shi, Y., Segal, E., and Chang, H.Y. (2010). Long noncoding RNA as modular scaffold of histone modification complexes. Science *329*, 689-693.

Valadkhan, S. (2005). snRNAs as the catalysts of pre-mRNA splicing. Curr Opin Chem Biol *9*, 603-608.

Vance, K.W., and Ponting, C.P. (2014). Transcriptional regulatory functions of nuclear long noncoding RNAs. Trends Genet *30*, 348-355.

Voigt, P., LeRoy, G., Drury, W.J., 3rd, Zee, B.M., Son, J., Beck, D.B., Young, N.L., Garcia, B.A., and Reinberg, D. (2012). Asymmetrically modified nucleosomes. Cell *151*, 181-193.

Voigt, P., Tee, W.W., and Reinberg, D. (2013). A double take on bivalent promoters. Genes Dev *27*, 1318-1338.

Walstrom, K.M., Schmidt, D., Bean, C.J., and Kelly, W.G. (2005). RNA helicase A is important for germline transcriptional control, proliferation, and meiosis in C. elegans. Mech Dev *122*, 707-720.

Wang, J., Rao, S., Chu, J., Shen, X., Levasseur, D.N., Theunissen, T.W., and Orkin, S.H. (2006). A protein interaction network for pluripotency of embryonic stem cells. Nature *444*, 364-368.

Wang, K.C., and Chang, H.Y. (2011). Molecular mechanisms of long noncoding RNAs. Mol Cell *43*, 904-914.

Wang, Y., Xu, Z., Jiang, J., Xu, C., Kang, J., Xiao, L., Wu, M., Xiong, J., Guo, X., and Liu, H. (2013). Endogenous miRNA sponge lincRNA-RoR regulates Oct4, Nanog, and Sox2 in human embryonic stem cell self-renewal. Dev Cell *25*, 69-80.

Wei, X., Pacyna-Gengelbach, M., Schluns, K., An, Q., Gao, Y., Cheng, S., and Petersen, I. (2004). Analysis of the RNA helicase A gene in human lung cancer. Oncol Rep *11*, 253-258.

Wen, B., Wu, H., Shinkai, Y., Irizarry, R.A., and Feinberg, A.P. (2009). Large histone H3 lysine 9 dimethylated chromatin blocks distinguish differentiated from embryonic stem cells. Nat Genet *41*, 246-250.

Wiblin, A.E., Cui, W., Clark, A.J., and Bickmore, W.A. (2005). Distinctive nuclear organisation of centromeres and regions involved in pluripotency in human embryonic stem cells. J Cell Sci *118*, 3861-3868.

Williams, R.L., Hilton, D.J., Pease, S., Willson, T.A., Stewart, C.L., Gearing, D.P., Wagner, E.F., Metcalf, D., Nicola, N.A., and Gough, N.M. (1988). Myeloid leukaemia inhibitory factor maintains the developmental potential of embryonic stem cells. Nature *336*, 684-687.

Wilmut, I., Schnieke, A.E., McWhir, J., Kind, A.J., and Campbell, K.H. (1997). Viable offspring derived from fetal and adult mammalian cells. Nature *385*, 810-813.

Wongtrakoongate, P., Riddick, G., Fucharoen, S., and Felsenfeld, G. (2015). Association of the Long Non-coding RNA Steroid Receptor RNA Activator (SRA) with TrxG and PRC2 Complexes. PLoS Genet *11*, e1005615.

Woodcock, C.L., Frado, L.L., and Rattner, J.B. (1984). The higher-order structure of chromatin: evidence for a helical ribbon arrangement. J Cell Biol *99*, 42-52.

Wray, J., Kalkan, T., and Smith, A.G. (2010). The ground state of pluripotency. Biochem Soc Trans *38*, 1027-1032.

Wu, S.C., and Zhang, Y. (2010). Active DNA demethylation: many roads lead to Rome. Nat Rev Mol Cell Biol *11*, 607-620.

Wutz, A., Rasmussen, T.P., and Jaenisch, R. (2002). Chromosomal silencing and localization are mediated by different domains of Xist RNA. Nat Genet *30*, 167-174.

Yang, F., Deng, X., Ma, W., Berletch, J.B., Rabaia, N., Wei, G., Moore, J.M., Filippova, G.N., Xu, J., Liu, Y., *et al.* (2015). The lncRNA Firre anchors the inactive X chromosome to the nucleolus by binding CTCF and maintains H3K27me3 methylation. Genome Biol *16*, 52.

Yang, L., Froberg, J.E., and Lee, J.T. (2014). Long noncoding RNAs: fresh perspectives into the RNA world. Trends Biochem Sci *39*, 35-43.

Yang, Q., Fairman, M.E., and Jankowsky, E. (2007). DEAD-box-protein-assisted RNA structure conversion towards and against thermodynamic equilibrium values. J Mol Biol *368*, 1087-1100.

Yang, R., Kerschner, J.L., Gosalia, N., Neems, D., Gorsic, L.K., Safi, A., Crawford, G.E., Kosak, S.T., Leir, S.H., and Harris, A. (2016). Differential contribution of cis-regulatory elements to higher order chromatin structure and expression of the CFTR locus. Nucleic Acids Res *44*, 3082-3094.

Yao, H., Brick, K., Evrard, Y., Xiao, T., Camerini-Otero, R.D., and Felsenfeld, G. (2010). Mediation of CTCF transcriptional insulation by DEAD-box RNA-binding protein p68 and steroid receptor RNA activator SRA. Genes Dev *24*, 2543-2555.

Ying, Q.L., Wray, J., Nichols, J., Batlle-Morera, L., Doble, B., Woodgett, J., Cohen, P., and Smith, A. (2008). The ground state of embryonic stem cell self-renewal. Nature *453*, 519-523.

Yokochi, T., and Robertson, K.D. (2002). Preferential methylation of unmethylated DNA by Mammalian de novo DNA methyltransferase Dnmt3a. J Biol Chem *277*, 11735-11745.

Yoon, J.H., Abdelmohsen, K., Srikantan, S., Yang, X., Martindale, J.L., De, S., Huarte, M., Zhan, M., Becker, K.G., and Gorospe, M. (2012). LincRNA-p21 suppresses target mRNA translation. Mol Cell *47*, 648-655.

Zhang, S., and Grosse, F. (1994). Nuclear DNA helicase II unwinds both DNA and RNA. Biochemistry *33*, 3906-3912.

Zhang, S., and Grosse, F. (1997). Domain structure of human nuclear DNA helicase II (RNA helicase A). J Biol Chem *272*, 11487-11494.

Zhang, S., Herrmann, C., and Grosse, F. (1999). Nucleolar localization of murine nuclear DNA helicase II (RNA helicase A). J Cell Sci *112 (Pt 16)*, 2693-2703.

Zhang, S., Schlott, B., Gorlach, M., and Grosse, F. (2004). DNA-dependent protein kinase (DNA-PK) phosphorylates nuclear DNA helicase II/RNA helicase A and hnRNP proteins in an RNA-dependent manner. Nucleic Acids Res *32*, 1-10.

Zhao, J., Ohsumi, T.K., Kung, J.T., Ogawa, Y., Grau, D.J., Sarma, K., Song, J.J., Kingston, R.E., Borowsky, M., and Lee, J.T. (2010). Genome-wide identification of polycomb-associated RNAs by RIP-seq. Mol Cell *40*, 939-953.

Zhao, J., Sun, B.K., Erwin, J.A., Song, J.J., and Lee, J.T. (2008). Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. Science *322*, 750-756.

Zhou, B.R., Jiang, J., Feng, H., Ghirlando, R., Xiao, T.S., and Bai, Y. (2015). Structural Mechanisms of Nucleosome Recognition by Linker Histones. Mol Cell *59*, 628-638.

Zhou, Y., and Grummt, I. (2005). The PHD finger/bromodomain of NoRC interacts with acetylated histone H4K16 and is sufficient for rDNA silencing. Curr Biol *15*, 1434-1438.

Zhou, Y., Santoro, R., and Grummt, I. (2002). The chromatin remodeling complex NoRC targets HDAC1 to the ribosomal gene promoter and represses RNA polymerase I transcription. Embo J *21*, 4632-4640.

Zhou, Y., Schmitz, K.M., Mayer, C., Yuan, X., Akhtar, A., and Grummt, I. (2009). Reversible acetylation of the chromatin remodelling complex NoRC is required for non-coding RNA-dependent silencing. Nat Cell Biol *11*, 1010-1016.

Curriculum vitae

PERSONAL INFORMATION

Surname	Leone
Name	Sergio
Date of birth	June 24, 1984
Hometown	Partinico, Italy
Nationality	Italian

EDUCATION

2012 – present	Doctoral Thesis in Molecular Biology (Dr. sc. nat./Ph.D.) University of Zurich
2009 – 2012	Master studies (M.Sc.) in Biotechnology for Industry and Scientific Research (BIRS) Faculty of Math, Physics and Natural Sciences, University of Palermo, Italy
2010 - 2011	Erasmus scholarship Fachhochschule Nordwestschweitz (FHNW), Basel, Switzerland
2009 (April-October)	Trainee at the National Research Council (CNR), Palermo, Italy
2003-2009	Bachelor studies in Biology (B. Sc.) Faculty of Math, Physics and Natural Sciences, University of Palermo, Italy
1998-2003	High School Diploma Liceo Scientifico S.Savarino, Partinico, Italy

PUBBLICATION LIST

Sergio Leone, Dominik Bär, Coenraad Frederik Slabber & Raffaella Santoro. Establishment of heterochromatin at rRNA genes is required for embryonic stem cell

differentiation. (Submitted for publication)

Sergio Leone & Raffaella Santoro (2016). Challenges in the analysis of long noncoding RNA functionality. FEBS Letters *590*, 2342–2353

Natasa Savic, Dominik Bär, Sergio Leone, Sandra C. Frommel, Fabienne A. Weber, Eva Vollenweider, Elena Ferrari, Urs Ziegler, Andres Kaech, Olga Shakhova, Paolo Cinelli & Raffaella Santoro (2014). lncRNA maturation to initiate heterochromatin formation in the nucleolus is required for exit from pluripotency in ESCs. Cell Stem Cell *15*, 720-734

Acknowledgements

Many persons contributed directly or indirectly to this work and it is time for me to express my gratitude.

First of all I want to thank my thesis supervisor PD Dr. Raffaella Santoro for having shed a bright light on the dark path of this PhD project with her scientific support, her courageous suggestions, her limitless motivation and her contagious passion for science.

I am very grateful to Prof. Dr. Dr. Michael O. Hottiger for helpful discussions related to my project and for being the chairman of my doctoral thesis committee. I want also to express my gratitude to the other members of my Thesis Committee, Prof. Dr. Ulrike Kutay, Prof. Dr. Christian Lehner, Dr. Davide Gabellini for their unconditioned help and support through the many constructive criticisms given during our meetings.

Many thanks to the former components of our research group Eva, Natasa and Sandra especially for their constant encouragement and their unlimited support during the first years of this experience, which kept me going on through the many difficult moments. I want also to thank them together with Julia, Anita and Nina for the countless funny and pleasant moments inside and outside the lab.

I want to warmly thank also the current members of Santoro's group, Damian, Sarah, Rodrigo, Karolina, Isabella and Dominik, for the friendly atmosphere they create in the lab, for the many funny moments, for the several well-timed coffee breaks, but also for the countless scientific and technical discussions that allowed us to deal with troubleshooting every day as a real team. A special thanks goes to my young "padawan" Erik for his contribution to this work, for giving me the opportunity to learn the difficult task of teaching as his lab supervisor and above all for making my final year of PhD the most funny and enjoyable.

It would be impossible to thank appropriately each former and current member of the DMMD and the CABMM for the exceptional work environment they generated and for the many useful and pleasant discussions. In this regard a special thanks goes to Dr. Ali Mirsaidi for his unstoppable conversation ability and for his supernatural gift of belittling every problem and cheer you up in any condition.

Many thanks to my dearest friends down in Sicily that make me always feel like I never left and to my friends here in Zürich that make me aware of a life outside the lab.

Last but not least I want to deeply thank my family for the extraordinary support throughout my whole education, for having always trusted my capability to reach this goal and for all the love they unselfishly give me.